Chartered Institute of Ergonomics & Human Factors

Contemporary Ergonomics & Human Factors 2025

Editors: Dave Golightly, Nora Balfe & Rebecca Charles

Contemporary Ergonomics & Human Factors 2025

Editors

Dave Golightly, Nora Balfe & Rebecca Charles

Chartered Institute of Ergonomics & Human Factors

ISBN: 978-1-9996527-7-7

Preface

Contemporary Ergonomics & Human Factors 2025 contains the proceedings of the Chartered Institute of Ergonomics & Human Factors Annual Conference, which was held 28-30 April 2025 at St George's Park, Burton upon Trent, UK.

Authors were invited to submit short, two-page papers covering the state of the art in human factors and ergonomics and we were delighted to again see an increase of in submissions, with 159 papers submitted this year. The papers were all peer-reviewed by at least three reviewers, and the Programme Committee met in January 2025 to make the final selection of short papers, long papers or posters for the conference. The accepted papers form these proceedings.

We continue to receive extremely high-quality submissions and are particularly pleased at the mix of academic or theoretical papers with more practice-based and reflective papers, which we feel well represents the UK ergonomics and human factors community.

As always, the conference represents a diverse range of application areas across healthcare, transport, energy, defence, construction and more. This year we are particularly pleased to include a special track from the medical device and combination products sector.

Themes this year range across human factors methods and tools, incident investigation, work design, technology and fatigue. AI continues to be a topic of increasing interest to the human factors community and that interest is well represented in this year's conference.

We are tremendously grateful for the contribution of our authors, our reviewers and of the Programme Committee who have helped to shape this year's proceedings.

Conference Programme Co-Chairs
Dave Golightly, Nora Balfe & Rebecca Charles

Annual Conference 2025 Programme Committee

Dave Golightly (Co-Chair)
Nora Balfe (Co-Chair)
Rebecca Charles (Co-Chair)
David Caple
Dipak Chauhan
Fay Dixon
Enda Fallon
Richard Farry
Alan Felstead
Ella-Mae Hubbard
Katie Plant
Andrew Thatcher

CIEHF Organisers

Tina Worthy, Chief Operating Officer
Ellie Poole, Communications & Content Manager

Contents

Construction

Cross sector

Defence

Emergency services

Energy

Healthcare

Manufacturing

Medical devices

Nuclear

Rail

Climate Change and Climate Action in Welsh Agriculture

Henry Amery

Cardiff University

SUMMARY

This paper explores the opinions and perspectives of Welsh farmers in relation to climate change, climate action and the future. In conjunction with, this paper explores how these opinions and perspectives translate into real life behaviours, and how the contextualisation of an 'environment' in terms of how abstract or how closely aligned it is to a farmer can impact their drive for climate action.

KEYWORDS

Climate ergonomics, agriculture, climate action, future thinking.

Introduction

There is an identifiable need for all sectors within Wales to do more to curb their contributions to climate change if Wales is to achieve its target of net zero emissions by 2050. The agricultural sector in particular needs to adapt its practices with current predictions suggesting that the sector could become the largest contributor to Welsh emissions by 2035 (Wales Centre for Public Policy, 2023), particularly as there is a recognised lack of future planning in relation to climate change within small businesses (SME Climate Hub, 2022; Mittal, 2024). Encouraging episodic future thinking (EFT), or the mental simulation of future events, could be useful in encouraging people to adopt more pro-environmental behaviours and climate action (Lee et al., 2020) which could in turn help with encouraging farmers to adjust the management of their farms. There is research ((Kuehne, 2012; Woods et al., 2017; Brobakk, 2018; Davidson et al., 2019; Houser, Gunderson and Stuart, 2019) to suggest that farmers have a lower than average acceptance of climate change, or have higher than average scepticism of it, with financial incentive or increased productivity acting as better motivators for climate action than the risk of adverse effects due to climate change itself. As such, with climate change risk perception acting as a mediator between EFT and climate conscious attitudes (Lee et al., 2020), farmers may find it more difficult to engage with EFT, but could still benefit nonetheless, particularly with the cumulative effect repeated EFT can have on delay discounting (Mellis et al., 2019). This study aimed to explore the opinions and perspectives of climate change and climate action, as well as the future thinking of Welsh farmers in order to contribute to the wider literature around the efficacy of EFT as an intervention in behaviours relating to climate action, as well as amplifying the opinions and concerns of farmers, an underrepresented group that have a huge impact on the natural environment. Furthermore, this study investigated the ways in which Welsh farmers engage with their own local environment, and how the way they view, and engage with, their environment affects the way they perceive climate change.

Methods

Semi-structured interviews were conducted to capture detailed perspectives, allowing the participants to share their opinions and perspectives, whilst affording the researcher the flexibility to explore interesting tangents whilst sticking to a pre-planned topic guide. Said topic guide structured the interviews into four sections: general farm information, opinions on climate change, future farming perspectives, and the interplay of future planning and climate change. Questions were open-ended to encourage detailed responses, supplemented by targeted probes where needed. A purposive sample of 10 Welsh farmers was chosen to reflect the proportional distribution of farm types in Wales. Farms were categorized by primary produce, with mixed farms proportionally divided. The inclusion criteria required participants to be principal farmers operating in Wales. Diversification enterprises were noted but not included in sample criteria. Interviews were conducted in person, recorded, and transcribed verbatim with identifying information removed. Thematic analysis was employed inductively, ensuring themes emerged organically from the data while acknowledging some deductive influence from the research questions.

Findings

Six core themes were identified from the interview data, with strong inter-theme relationships established. These themes included Climate Change, Climate Action, Government Policy, Barriers to Climate Action, Future of Farming and Future Thinking. Furthermore, these themes had strong inter-theme relationships which are described in table 1.

Table 1: Inter-theme relationships

Inter-theme relationship	Description
Climate Change & Barriers to Climate Action	Participants who were sceptical about climate change often viewed climate action as futile, citing it as a waste of time or resources.
Government Policy & Barriers to Climate Action	Distrust in government policy aligned with scepticism about climate action, particularly regarding political motives and practical challenges.
Climate Action & Future Thinking	Participants actively engaged in climate action often emphasized its economic benefits but also acknowledged environmental advantages, linking sustainability with clearer visions of the future.
Barriers to Climate Action & Future of Farming	Financial and systemic barriers to climate action were linked to concerns about the wider farming sector's insecure future.
Climate Change & Climate Action	Despite scepticism about climate change when talking about the environment at large, participants showed strong commitment to preserving their farmland, revealing a personal connection to environmental stewardship. Optimism was higher when discussing the participant's personal futures rather than the wider sector.

Perhaps the most interesting finding to emerge was the stark difference in attitudes participants had when talking about different kinds of environments. When talking about the more abstract (to them) wider environment, they appeared to believe they had little to no responsibility to protect it, or that climate change even exists. However, when the environment being discussed was their own local environment, there was a real sense of duty to protect it against changing and increasingly adverse weather. This suggests that the context with which farmers are approached in relation to climate action and sustainable practices can have a significant impact on their acceptance of pro-climate measures. However, this intense sense of duty for their local environments appears to pose challenges for policy makers/governments who were widely considered to be an untrustworthy outgroup with questionable motives.

References

Brobakk, J. (2018). A climate for change? Norwegian farmers' attitudes to climate change and climate policy. *World Political Science*, *14*(1), 55-79. https://doi.org/10.1515/wps-2018-0003

Davidson, D. J., Rollins, C., Lefsrud, L., Anders, S., & Hamann, A. (2019). Just don't call it climate change: climate-skeptic farmer adoption of climate-mitigative practices. *Environmental Research Letters*, *14*(3), 034015. https://doi.org/10.1088/1748-9326/aafa30

Houser, M., Gunderson, R., & Stuart, D. (2019). Farmers' perceptions of climate change in context: Toward a political economy of relevance. *Sociologia Ruralis*, *59*(4), 789-809. https://doi.org/10.1111/soru.12268

Kuehne, G. (2012). How do farmers' climate change beliefs affect adaptation to climate change?. *Society & Natural Resources*, *27*(5), 492-506. https://doi.org/10.1080/08941920.2013.861565

Lee, P. S., Sung, Y. H., Wu, C. C., Ho, L. C., & Chiou, W. B. (2020). Using episodic future thinking to pre-experience climate change increases pro-environmental behavior. *Environment and Behavior*, *52*(1), 60-81. https://doi.org/10.1177/0013916518790590

Mellis, A. M., Snider, S. E., Deshpande, H. U., LaConte, S. M., & Bickel, W. K. (2019). Practicing prospection promotes patience: Repeated episodic future thinking cumulatively reduces delay discounting. *Drug and alcohol dependence*, *204*, 107507. https://doi.org/10.1016/j.drugalcdep.2019.06.010

Mittal, S. (2024). *Survey: Two-thirds of SMEs struggling to cut carbon due to financial hurdles.* Edie. https://www.edie.net/survey-two-thirds-of-smes-struggling-to-cut-carbon-due-to-financial-hurdles/#:~:text=New%20research%20has%20found%20that,cited%20as%20a%20significant%20obstacle.

SME Climate Hub. (2022, February). *New data reveals two-thirds of surveyed small businesses concerned over navigating climate action.* SME Climate Hub. https://businessclimatehub.org/new-survey-reveals-small-business-barriers-climate-action/

Wales Centre for Public Policy. (2023). *How could Wales feed itself in 2035? Evidence Pack.* Wales Centre for Public Policy. https://www.wcpp.org.uk/wp-content/uploads/2023/07/How-could-Wales-feed-itself-in-2035-Evidence-Pack-1.pdf

Woods, B. A., Nielsen, H. Ø., Pedersen, A. B., & Kristofersson, D. (2017). Farmers' perceptions of climate change and their likely responses in Danish agriculture. *Land use policy*, *65*, 109-120. https://doi.org/10.1016/j.landusepol.2017.04.007

Head Movements and Physiological Indicators as Predictors of Passenger Motion Sickness

William Emond[1,2] & Mohsen Zare [2]

[1]Mercedes-Benz AG, Germany, [2]ELLIADD-ERCOS, UTBM, France

SUMMARY

Future car transportation is expected to feature automated driving. In such vehicles, drivers will no longer seat behind the wheel but evolve into passengers who may focus on non-driving related activities during the drive. This figure is however expected to increase the incidence and severity of motion sickness while traveling, which motivates the need of an improved understanding of passenger motion sickness and the design of comprehensive mitigation solutions. This research investigated the relationship between passenger motion sickness, physiological parameters and postural activity reflected in head dynamics. In an 18-minute realistic passenger drive scenario, 12 (1 female, 11 male) participants were tested in two separate sessions. Their physiological parameters were recorded using a medical device and their head movements were recorded using motion capturing. Using a Generalized Linear Model, analyses identified changes in peripheral oxygen saturation levels, core temperature and cardiovascular activity as physiological reactions with strong relationships to motion sickness severity. Moreover, the amplitude of head movements in the roll and yaw directions showed significant relationships with motion sickness severity. These findings pave the road for an improved detection of motion sickness depending on passenger individual parameters.

KEYWORDS

Kinetosis, Car Sickness, Human Factors, Motion Comfort, Non-Driving Task

Introduction

Road vehicles are currently undergoing a revolution in their design and use. Future cars are forecasted to feature automated driving with the possibility for the user to quit the driving task and dedicate travel time to Non-Driving Related Activities (NDRA). Passenger comfort will become a key requirement for the design of such vehicles as moving living spaces (Salter et al., 2019). Taking users "out of the loop" for driving the vehicle may result in a greater likeliness to spend travel time on NDRAs. However, focusing on NDRAs devoid of any view of the vehicle surroundings increases the likeliness of experiencing motion sickness (Diels, 2014; Metzulat et al., 2024; Morimoto et al., 2008). Other dimensions inherent to automated vehicles such as the (un)familiarity of the driving style and (lack of) trust in automation may even exacerbate this prevalence (Peng et al., 2024). Experts forecast that, with the introduction of self-driving cars, the occurrence of motion sickness may nearly double in road transportation (Diels et al., 2016; Iskander et al., 2019). A figure which may jeopardize the successful introduction of such vehicles on public roads. Therefore, recent years witnessed consequent research on the design of systems capable of alleviating motion sickness while traveling in a car. This research is mostly driven by automotive manufacturers who target to be the first to market for successfully deploying self-driving cars on the road (Diels, 2014).

Motivation and Objective

For a comprehensive mitigation of motion sickness, there is also a need towards the development of methods for predicting the time course of symptoms. An optimal countermeasure of motion sickness may be able to trigger efficient alleviation methods as soon as a rise in motion sickness severity is increased. Considering the high inter-individual variability in autonomic reactions to motion sickness and wide spectrum of sensitivities, such detection methods should be user-centred and involve the measurement of several parameters. Literature emphasized the combination of physiological measurements as a reliable prediction method (Tan et al., 2022). Furthermore, postural activity, and notably the measurement of head rotations, was documented as another predictor parameter of motion sickness (Wada et al., 2012).

Therefore, two research questions were investigated:

RQ1. How different head movements variables correlate with motion sickness severity?

RQ2. How do changes in physiological parameters correlate with the onset and severity of passenger motion sickness?

The present study addressed these research questions by replicating a realistic passenger drive reproducing fore-aft accelerations, representative of an everyday passenger commute.

Methodology

Twelve (11 male, 1 female) volunteers experienced an 18-minute vehicle drive on the passenger seat of a saloon vehicle (Mercedes-Benz E 400 e) driven six identical laps on a closed test track. Their mean ± SD age was 41 ± 12 years and their mean ± SD motion sickness susceptibility score as assessed by the short version of the Motion Sickness Susceptibility Questionnaire (Golding, 2006) was 17.7 ± 9.5. In a within-subjects design, they tested either an "Upright" or "Reclined" sitting condition in randomized order: in the Upright condition, the backrest position was kept fixed at 28 degrees from the vertical, whereas in the Reclined condition, participants were free to choose a backrest angle superior to 28 degrees, resulting in a mean ± SD angle inclination of 33 ± 3 degrees from the vertical. For each participant, the two experimental sessions were spaced by at least 24 hours.

During the experimental drive, subjective motion sickness ratings were collected each minute. The order of conditions was not counterbalanced (3 participants were first tested in the Upright condition). Motion sickness severity was assessed every minute (three times per lap) using a 21-point "Fast Motion Sickness (FMS)" scale (0 being very good, 20 being strong nausea) referring to individual wellbeing (Keshavarz & Hecht, 2011). Participants were free in their gazing behaviour and could withdraw from the experimental sessions at any time. Physiological data of participants was monitored and recorded using an in-ear medical device (Cosinuss c-med° alpha) performing continuous measurements of heart rate, RR-intervals at 1 Hz, as well as measurements of core temperature, oxygen saturation level, and perfusion index at 0.1 Hz. An additional quality index signal was generated to help in selecting reliable data for the data analysis. Complementary to physiological measurements, head movements were recorded during the drive using motion capturing techniques: a camera (OptiTrack V:120 Duo) was fixed to the windshield and facing the participants at eye-height. To track head movements, four targets were placed on the face of the participant (one on the upper forehead, one between the eyebrows and one on each zygoma extremity). Both the camera and the in-ear devices were calibrated with participants sitting still in the vehicle before starting the experimental drive.

For data analysis, head movements and physiological measurements were averaged on each time interval between the successive FMS scores of participants. Heart Rate Variability and Low-

Frequency/High-Frequency (LF/HF) ratio were derived from RR-intervals. Using the Matlab (MathWorks, version R2024a) software, the Root Mean Square (RMS), skewness (skw) and kurtosis (kts) of head movement data were calculated in the three directions of rotation and rotation velocity: namely pitch (x-axis) for neck flexion/extension, yaw (y-axis) for head rotation and roll (z-axis) for lateral bending. To attenuate vibration noise related to the drive, signals were processed with a Butterworth filter with cutoff frequency set at 5 Hz and zero-phase filtering. Statistical analyses were performed with R Studio (Version 4.3.3, Posit) with significance level set at $p < .05$. Since FMS scores were not normally distributed, an analysis of FMS scores using a Generalized Linear Model (GLM) was performed.

Results

The free view on the vehicle surroundings and the absence of any restriction to focus on a NDRA resulted in moderate motion sickness severity on average (Figure 1). Seven participants experienced some motion sickness symptoms in both experimental sessions and one male participant did not feel any symptom in both sessions.

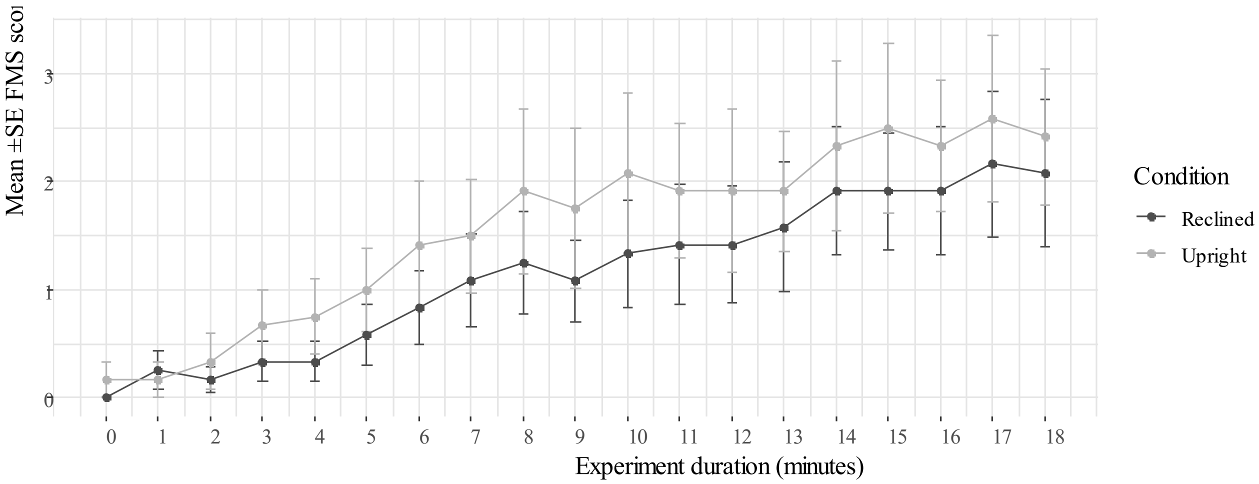


Figure 1: Time course of FMS scores in each experimental condition, n = 12 participants, within-subjects design.

The GLM analysis (Table 1) revealed that the duration of the drive and the individual susceptibility (measured with the MSSQ-Short) had strong relationship with motion sickness severity. No significant effect of the backrest angle could be identified (t = 0.747, p = .456). However, several physiological parameters were identified to correlate with the time course of FMS scores, notably increases in perfusion index (t = 6.139, p < .001), decreases in peripheral oxygen saturation (t = -4.217, p < .001), decreases in core body temperature (t = -3.317, p = .001) and increases in heart rate (t = 2.423, p = .016). Heart Rate Variability and LF/HF ratio had no significant relationship with FMS scores.

The GLM analysis of FMS scores identified two characteristics of head dynamics as related to increases in motion sickness severity: namely increases in the amplitude of head roll rotations (t = 3.442, p < .001) and decreases of the amplitude of head yaw angle rotations (t = -3.201, p = .002). No effect of skewness or kurtosis of head rotation could be identified. Additionally, the analysis did not identify any relationship between FMS scores and characteristics of the head rotation velocity.

Table 1: Results from the Generalized Linear Model on the time course of FMS scores, n = 12 participants.

	Estimate	Std. Error	t value	p value	odds ratio*	VIF**
(Intercept)	6.82e+01	1.26e+01	5.405	2.24e-07	4.27e+29	-
Drive duration (min)	1.42e-01	2.19e-02	6.499	9.15e-10	1.15e+00	1.60e+00
MSSQ score (-)	1.77e-01	1.89e-02	9.360	5.28e-17	1.19e+00	2.98e+00
Backrest angle (deg)	3.32e-02	4.45e-02	0.747	4.56e-01	1.03e+00	1.89e+00
Heart rate (min^{-1})	2.77e-02	1.15e-02	2.423	1.65e-02	1.03e+00	3.89e+00
Core temperature (°C)	-1.34e+00	4.03e-01	-3.317	1.12e-03	7.57e-01	5.12e+00
Peripheral oxygen saturation (%)	-2.78e-01	6.60e-02	-4.217	4.07e-05	2.62e-01	2.21e+00
Perfusion index (-)	2.55e+00	4.15e-01	6.139	5.96e-09	1.28e+01	2.58e+00
Heart Rate Variability (ms)	-7.19e-05	1.54e-03	-0.047	9.63e-01	1.00e+00	1.49e+00
LF/HF ratio (-)	4.14e-02	7.77e-02	0.533	5.95e-01	1.04e+00	1.38e+00
RMS pitch angle (deg)	1.90e-02	1.73e-02	1.099	2.73e-01	1.02e+00	1.83e+00
RMS yaw angle (deg)	-1.22e-01	3.82e-02	-3.201	1.64e-03	8.85e-01	1.78e+00
RMS roll angle (deg)	5.00e-02	1.45e-02	3.442	7.31e-04	1.05e+00	2.02e+00
Skewness of pitch angle (-)	-8.39e-02	1.41e-01	-0.595	5.52e-01	9.20e-01	2.15e+00
Skewness of yaw angle (-)	3.14e-01	1.73e-01	1.814	7.15e-02	1.37e+00	1.43e+00
Skewness of roll angle (-)	6.66e-02	1.05e-01	0.637	5.25e-01	1.07e+00	1.43e+00
Kurtosis of pitch angle (-)	-2.20e-02	3.90e-02	-0.565	5.73e-01	9.78e-01	2.27e+00
Kurtosis of yaw angle (-)	-1.45e-02	8.48e-02	-0.171	8.64e-01	9.86e-01	1.75e+00
Kurtosis of roll angle (-)	-1.40e-02	3.20e-02	-0.437	6.63e-01	9.86e-01	1.88e+00
RMS pitch angle velocity ($deg.s^{-1}$)	-6.27e-01	1.69e+00	-0.372	7.11e-01	5.34e-01	3.39e+00
RMS yaw angle velocity ($deg.s^{-1}$)	2.25e-01	1.74e+00	0.129	8.97e-01	1.25e+00	1.33e+00
RMS roll angle velocity ($deg.s^{-1}$)	5.39e-01	5.43e-01	-0.993	3.22e-01	5.83e-01	3.48e+00
Skewness of pitch angle velocity (-)	7.48e-01	1.17e-01	0.638	5.25e-01	1.08e+00	2.03e+00
Skewness of yaw angle velocity (-)	1.30e-02	1.86e-01	0.699	4.85e-01	1.14e+00	1.59e+00
Skewness of roll angle velocity (-)	9.44e-02	1.15e-01	0.820	4.14e-01	1.10e+00	1.84e+00
Kurtosis of pitch angle velocity (-)	1.17e-02	1.15e-02	1.019	3.10e-01	1.01e+00	3.10e+00
Kurtosis of yaw angle velocity (-)	6.18e-03	2.70e-02	0.229	8.19e-01	1.01e+00	1.99e+00
Kurtosis of roll angle velocity (-)	-8.45e-03	1.86e-02	-0.454	6.50e-01	9.92e-01	3.02e+00

* Odds ratio quantifies the change in odds of the dependent event occurring for a one-unit increase in the predictor variable

** Variance Inflation factor (VIF) is a measure of the extent of multicollinearity in the set of multiple regression variables

Discussion

Results from motion capturing data support an influence of postural activity on the development of passenger motion sickness. The contributing effect of head roll movements to the severity of symptoms aligns with observations from other studies replicated in vehicular environments (Papaioannou et al., 2024; Wada et al., 2012; Wada & Yoshida, 2016). However, the inversely proportional effect of head yaw amplitude of rotation is quite novel and might reflect an orientation behaviour while gazing at specific visual cues from the vehicle surroundings. This attitude might also reflect distraction, which is reported as a protective behaviour against the development of symptoms (Bos, 2015). Referring to the first research question (RQ1), head movements stood out as correlates of motion sickness severity and could be considered for the development of camera-based motion sickness prediction models in vehicles.

The significant relationship between perfusion index and FMS scores suggests that a rise in motion sickness severity would be accompanied by an increase in blood pressure or volume. This finding aligns with similar observations of an increased blood volume as a reaction to passenger motion sickness (Pham Xuan et al., 2019). Yet, other findings from empirical research suggest no relationship (Graybiel & Lackner, 1980) or a negative relationship between blood pressure and motion sickness severity (Steele, 1968; Tan et al., 2022). This increase in cardiovascular activity is also reflected into a slight heart rate increase correlating with motion sickness, which aligns with findings from motion sickness studies (Cowings et al., 1986; Henry et al., 2023; Irmak et al., 2021).

No experimental research investigated oxygen saturation levels respective to motion sickness, the negative relationship identified aligns with findings from genetic research identifying a correlation between motion sickness susceptibility and sensitivity to hypoxia (Hromatka et al., 2015). Whether this body reaction is a cause or a consequence of motion sickness is however unclear. A physiological reaction consequent to motion sickness is hypothermia (Nalivaiko et al., 2014; Nobel et al., 2006), which could be identified by the negative relation ship between FMS scores and core temperature. Similar findings were reported in a similar stop-and-go passenger ride experiment (Pham Xuan et al., 2019). Referring to the second research question (RQ2), the significant relationships identified between many physiological parameters and FMS scores support the development of motion sickness detection methods based on the real-time monitoring of such parameters.

The relatively low duration of the passenger ride, the restriction of manoeuvres to fore-aft accelerations, and the limited sample size constitute strong limitations to the generalizability of findings. Replicating a design to elicit more severe forms of motion sickness (e.g restricting forward vision of participants) might yield greater differences in FMS scores across participants for a more generalizable analysis but would limit the real representativeness of the drive scenario. Increasing the duration of testing may be judged by volunteering participants as too time-consuming. The sample was also biased by unequal gender and age repartition, which are not representative of a general population. Further research should consider samples that are more representative of car passengers, with all ages, gender and susceptibilities represented to better understand the effect of these independent variables.

Conclusion

The present study aimed investigating the development of motion sickness over a duration of 18 minutes in a realistic passenger drive scenario replicating fore-aft vehicle accelerations. Despite moderate motion sickness severity, some significant correlation could be identified. Oxygen saturation, core temperature and cardiovascular activity stood out as indicators of a rise in cardinal motion sickness symptoms. Moreover, the significant relationships identified between motion sickness severity and head movements suggest a contributing effect of postural activity while riding as a vehicle passenger with free gazing behaviour. Despite limited generalizability, these findings can guide the development of systems capable of early detection of motion sickness to prevent the symptoms from exacerbating.

Acknowledgement

The authors would like to thank the participants of the study and are pleased to acknowledge Josef Ghebru (University of Stuttgart) as second experimenter.

References

Bohrmann, D., Bruder, A., & Bengler, K. (2022). Effects of Dynamic Visual Stimuli on the Development of Carsickness in Real Driving. *IEEE Transactions on Intelligent Transportation Systems*, 1–10. IEEE Transactions on Intelligent Transportation Systems. https://doi.org/10.1109/TITS.2021.3128834

Bohrmann, D., Koch, T., Maier, C., Just, W., & Bengler, K. (2020). *Motion Comfort – Human Factors of Automated Driving*. 1697–1708. https://www.researchgate.net/profile/Dominique-Bohrmann/publication/346647590_Motion_Comfort_-_Human_Factors_of_Automated_Driving/links/5fcc0619299bf188d4f96833/Motion-Comfort-Human-Factors-of-Automated-Driving.pdf

Bos, J. E. (2015). Less sickness with more motion and/or mental distraction. *Journal of Vestibular Research*, *25*(1), 23–33. https://doi.org/10.3233/VES-150541

Cowings, P. S., Suter, S., Toscano, W. B., Kamiya, J., & Naifeh, K. (1986). General Autonomic Components of Motion Sickness. *Psychophysiology*, *23*(5), 542–551. https://doi.org/10.1111/j.1469-8986.1986.tb00671.x

D'Amour, S., Bos, J. E., & Keshavarz, B. (2017). The efficacy of airflow and seat vibration on reducing visually induced motion sickness. *Experimental Brain Research*, *235*(9), Article 9. https://doi.org/10.1007/s00221-017-5009-1

Diels, C. (2014). Will autonomous vehicles make us sick? In S. Sharples & S. Shorrock (Eds.), *Contemporary Ergonomics and Human Factors 2014* (pp. 301–307). Taylor and Francis. https://doi.org/10.13140/RG.2.1.1461.0087

Diels, C., & Bos, J. (2021). Great Expectations: On the Design of Predictive Motion Cues to Alleviate Carsickness. In H. Krömker (Ed.), *HCI in Mobility, Transport, and Automotive Systems* (Vol. 12791, pp. 240–251). Springer International Publishing. https://doi.org/10.1007/978-3-030-78358-7_16

Diels, C., Bos, J. E., Hottelart, K., & Reilhac, P. (2016). Motion Sickness in Automated Vehicles: The Elephant in the Room. In G. Meyer & S. Beiker (Eds.), *Road Vehicle Automation 3* (pp. 121–129). Springer International Publishing. https://doi.org/10.1007/978-3-319-40503-2_10

Emond, W., & Zare, M. (2024). *Motion Sickness Mitigation in Road Vehicles with a Triad of Focus on Safety and Sustainability.* 8. https://hal.science/hal-04538188/document

Golding, J. F. (2006). Predicting individual differences in motion sickness susceptibility by questionnaire. *Personality and Individual Differences*, *41*(2), Article 2. https://doi.org/10.1016/j.paid.2006.01.012

Graybiel, A., & Lackner, J. R. (1980). Evaluation of the relationship between motion sickness symptomatology and blood pressure, heart rate, and body temperature. *Aviation, Space, and Environmental Medicine*, *51*(3), 211–214.

Henry, E. H., Bougard, C., Bourdin, C., & Bringoux, L. (2023). Car sickness in real driving conditions: Effect of lateral acceleration and predictability reflected by physiological changes. *Transportation Research Part F: Traffic Psychology and Behaviour*, *97*, 123–139. https://doi.org/10.1016/j.trf.2023.06.018

Hromatka, B. S., Tung, J. Y., Kiefer, A. K., Do, C. B., Hinds, D. A., & Eriksson, N. (2015). Genetic variants associated with motion sickness point to roles for inner ear development, neurological processes and glucose homeostasis. *Human Molecular Genetics*, *24*(9), 2700–2708. https://doi.org/10.1093/hmg/ddv028

Htike, Z., Papaioannou, G., Siampis, E., Velenis, E., & Longo, S. (2022). Fundamentals of Motion Planning for Mitigating Motion Sickness in Automated Vehicles. *IEEE Transactions on Vehicular Technology*, *71*(3), 2375–2384. IEEE Transactions on Vehicular Technology. https://doi.org/10.1109/TVT.2021.3138722

Irmak, T., Pool, D. M., & Happee, R. (2021). Objective and subjective responses to motion sickness: The group and the individual. *Experimental Brain Research*, *239*(2), 515–531. https://doi.org/10.1007/s00221-020-05986-6

Iskander, J., Attia, M., Saleh, K., Nahavandi, D., Abobakr, A., Mohamed, S., Asadi, H., Khosravi, A., Lim, C. P., & Hossny, M. (2019). From car sickness to autonomous car sickness: A review. *Transportation Research Part F: Traffic Psychology and Behaviour*, *62*, 716–726. https://doi.org/10.1016/j.trf.2019.02.020

Jurisch, M., Holzapfel, C., & Buck, C. (2020). The influence of active suspension systems on motion sickness of vehicle occupants. *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 1–6. https://doi.org/10.1109/ITSC45102.2020.9294311

Keshavarz, B., & Hecht, H. (2011). Validating an Efficient Method to Quantify Motion Sickness. *Human Factors*, *53*(4), Article 4. https://doi.org/10.1177/0018720811403736

Kuiper, O. X., Bos, J. E., Diels, C., & Schmidt, E. A. (2020). Knowing what's coming: Anticipatory audio cues can mitigate motion sickness. *Applied Ergonomics*, *85*, 103068. https://doi.org/10.1016/j.apergo.2020.103068

Metzulat, M., Metz, B., Landau, A., Neukum, A., & Kunde, W. (2024). Does the visual input matter? Influence of non-driving related tasks on car sickness in an open road setting. *Transportation Research Part F: Traffic Psychology and Behaviour*, *104*, 234–248. https://doi.org/10.1016/j.trf.2024.06.002

Morimoto, A., Isu, N., Asano, H., Kawai, A., & Masui, F. (2008). *Effects of reading books and watching movies on inducement of car sickness.* 14–19. https://www.researchgate.net/profile/Fumito-Masui/publication/264843228_Effects_of_Reading_Books_and_Watching_Movies_on_Inducement_of_Car_Sickness/links/542eb6fd0cf27e39fa963cb6/Effects-of-Reading-Books-and-Watching-Movies-on-Inducement-of-Car-Sickness.pdf

Nalivaiko, E., Rudd, J. A., & So, R. H. (2014). Motion sickness, nausea and thermoregulation: The "toxic" hypothesis. *Temperature*, *1*(3), Article 3. https://doi.org/10.4161/23328940.2014.982047

Nobel, G., Eiken, O., Tribukait, A., Kölegård, R., & Mekjavic, I. B. (2006). Motion sickness increases the risk of accidental hypothermia. *European Journal of Applied Physiology*, *98*(1), 48–55. https://doi.org/10.1007/s00421-006-0217-6

Papaioannou, G., Desai, R., & Happee, R. (2024). The Impact of Body and Head Dynamics on Motion Comfort Assessment. In W. Huang & M. Ahmadian (Eds.), *Advances in Dynamics of Vehicles on Roads and Tracks III* (pp. 54–63). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-66968-2_6

Peng, C., Horn, S., Madigan, R., Marberger, C., Lee, J. D., Krems, J., Beggiato, M., Romano, R., Chongfeng, W., Wooldridge, E., Happee, R., Hagenzieker, M., & Merat, N. (2024). Conceptualising user comfort in automated driving: Findings from an expert group workshop. *Transportation Research Interdisciplinary Perspectives*, *24*, 101070. https://doi.org/10.13140/RG.2.2.14206.87369

Pham Xuan, R., Brietzke, A., & Marker, S. (2019). *Evaluation of physiological responses due to car sickness with a zero-inflated regression approach*. 147–160. https://www.hfes-europe.org/wp-content/uploads/2019/10/PhamXuan2019.pdf

Salter, S., Diels, C., Herriotts, P., Kanarachos, S., & Thake, D. (2019). Motion sickness in automated vehicles with forward and rearward facing seating orientations. *Applied Ergonomics*, *78*, 54–61. https://doi.org/10.1016/j.apergo.2019.02.001

Schartmüller, C., & Riener, A. (2020). Sick of Scents: Investigating Non-invasive Olfactory Motion Sickness Mitigation in Automated Driving. *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 30–39. https://doi.org/10.1145/3409120.3410650

Steele, J. E. (1968). *The symptomatology of motion sickness* (AMRL-TR-70-86; Fourth Symposium on The Role of the Vestibular Organs in Space Exploration, pp. 89–96). Naval Aerospace Medical Institute, Pensacola, Florida.

Stelling, D., Hermes, M., Huelmann, G., Mittelstädt, J. M., Niedermeier, D., Schudlik, K., & Duda, H. (2021). Individual differences in the temporal progression of motion sickness and anxiety:

The role of passengers' trait anxiety and motion sickness history. *Ergonomics*, *64*(8), 1062–1071. https://doi.org/10.1080/00140139.2021.1886334

Tan, R., Li, W., Hu, F., Xiao, X., Li, S., Xing, Y., Wang, H., & Cao, D. (2022). Motion Sickness Detection for Intelligent Vehicles: A Wearable-Device-Based Approach. *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, 4355–4362. https://doi.org/10.1109/ITSC55140.2022.9922392

Wada, T., Konno, H., Fujisawa, S., & Doi, S. (2012). Can Passengers' Active Head Tilt Decrease the Severity of Carsickness?: Effect of Head Tilt on Severity of Motion Sickness in a Lateral Acceleration Environment. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *54*(2), 226–234. https://doi.org/10.1177/0018720812436584

Wada, T., & Yoshida, K. (2016). Effect of passengers' active head tilt and opening/closure of eyes on motion sickness in lateral acceleration environment of cars. *Ergonomics*, *59*(8), 1050–1059. https://doi.org/10.1080/00140139.2015.1109713

Yen Pik Sang, F. D., Billar, J. P., Golding, J. F., & Gresty, M. A. (2003). Behavioral Methods of Alleviating Motion Sickness: Effectiveness of Controlled Breathing and a Music Audiotape. *Journal of Travel Medicine*, *10*(2), 108–111. https://doi.org/10.2310/7060.2003.31768

Yusof, N. Md., Karjanto, J., Terken, J. M. B., Delbressine, F. L. M., & Rauterberg, G. W. M. (2020). Gaining Situation Awareness through a Vibrotactile Display to Mitigate Motion Sickness in Fully-Automated Driving Cars. *International Journal of Automotive and Mechanical Engineering*, *17*(1), 7771–7783. https://doi.org/10.15282/ijame.17.1.2020.23.0578

Developing Preliminary Heuristics for In-Vehicle Ambient Intelligence Systems

Shriya Pande[1], Catherine Harvey[2] & David R Large[2]

[1]School of Computer Science, [2]Human Factors Research Group, University of Nottingham

SUMMARY

This study aimed to develop heuristics for designing and evaluating in-vehicle Ambient Intelligence (AmI) systems. Utilising an expert focus group (n=3) and a user study (n=8), 15 preliminary heuristics were derived. These aim to ensure that in-vehicle AmI enhances the driving experience by fostering driver wellbeing and a positive user experience, whilst minimising distraction.

KEYWORDS

Ambient Intelligence, In-Vehicle, Driving, Heuristics

Introduction

Ambient Intelligence (AmI) describes, *"the application and embedding of artificial intelligence into everyday environments to seamlessly provide assistive and predictive support in a multitude of scenarios via an invisible user interface"* (Dunne et al., 2021). AmI systems have been integrated into many everyday settings, including homes, classrooms, hospitals etc., where they monitor and detect users' behaviour and intervene autonomously to enhance wellbeing, reduce cognitive load, increase engagement/enjoyment, etc. (Augusto et al., 2013). However, as the contexts of use become more complex, it is increasingly difficult to articulate users' needs and expectations and to determine how to address these (Stephanidis et al., 2021). Moreover, as technology develops, new experiences become possible, such as 'phygital' devices that bridge the digital and physical worlds.

In a driving context, there is also complexity in that AmI systems must enhance the in-vehicle user experience while not interrupting or disrupting the primary driving task and this may present a conflict. For example, evocative lighting or music may be used to enhance the driver's mood or their kinematic perception of the vehicle's performance but could inadvertently direct their attention away from the road situation or encourage riskier driving behaviour. Further complications arise because each experience should be individually tailored to the driver's preferences and lifestyle choices, which can change based on various factors, such as the journey type, presence of passengers etc. Nevertheless, recent examples suggest that the automotive industry is committed to creating user-centred 'intelligent' vehicle environments, with a focus on enhancing drivers' wellbeing, in particular. Concepts include the Audi Urbansphere, which uses facial recognition to determine the driver's stress level and reclines their seat to *"maximise comfort and restoration"*, and JLR's Body and Soul Seat (BASS), which is described as an *"AI tactile audio device"* embedded in the driver's seat that aims to enhance drivers' wellness using 'vibroacoustic therapy' (Fortune, 2024). However, these concepts generally lack user-centred testing and validation in a driving context, and it is therefore unclear whether the proposed AmI solutions deliver what they purport to without compromising the driving task or the driver's role within it. Traditional user experience and usability techniques, and indeed traditional 'driver distraction' guidelines, lack the scope or specificity to evaluate the nuances of in-vehicle AmI, and new or enhanced methods are subsequently required to define bespoke user requirements and appropriate evaluation metrics.

A common approach in user-centred design and evaluation is to apply ‘heuristics’ (or ‘rules-of-thumb’) to identify issues with a system or interface (Nielsen and Molich, 1990). Heuristic evaluation is a cost-effective method as it requires minimal training; by definition, heuristics should be easy to understand and apply, and the technique itself is intended to be intuitive and easy to accomplish. It is particularly effective as a formative, ‘discount’ method applied early in the design cycle, as it does not require systems to be complete or fully functional, but the method can also be applied as a summative evaluation of more mature systems. Nielsen’s ‘10 Usability Heuristics’ (Nielsen, 2024) have been applied in countless situations, but as they are targeted at direct interactions with a system, they are difficult to interpret and apply in the unique and nuanced context of in-vehicle AmI where only passive interaction is intended.

Overview of Study

The study aimed to develop domain-specific heuristics for in-vehicle AmI, with a particular focus on driver wellbeing. Firstly, a focus group was conducted with 3 experts in the field of automotive HF, HMI design and AmI. In addition to providing specific insights, this also informed the design of a user study that took place in a driving simulator, in which 8 participants (3M, 5F) experienced a simulated journey in which a prototypical AmI interface provided immersive sensory stimuli (dynamic images, sound, lighting) using Wizard-of-Oz, and were interviewed afterwards. Finally, inductive thematic analysis was conducted on all transcribed data to derive relevant themes and terms around which to define the heuristics. Fifteen heuristics were subsequently derived (Table 1).

Table 1: Preliminary AmI Heuristics

1. Stimuli must be balanced to prioritise safety and enhance the driving experience of occupants without causing distractions that hinder the primary task of driving. 2. The system should be subtle and non-explicit in nature by blending into the background to support the driving task without causing distractions. 3. The system should have situational awareness. (i.e. time, speed, traffic conditions, weather etc.). 4. The system output methods should be in sync with each other to complement the vehicle settings and user-state 5. The system should enhance the multi-sensory ambient interaction by incorporating any/all of visual, auditory, haptic and/or olfactory stimuli to create the in-vehicle environment. 6. A balance between comfort and alertness should be maintained in the vehicle. 7. The system should be adaptive in response to different users (i.e. offer personalisation without user input). 8. The system should be integrated with driver monitoring technology allowing it to adapt to occupant’s current state and intervene accordingly, whenever necessary. 9. System integration should complement driving information outputs (e.g. speed) rather than replace it, without overwhelming the occupant with unnecessary information.	10. The system should avoid complicated and over-saturated stimuli for extended periods of time to prevent mental overload and reduced driver alertness. 11. The system should preserve user privacy by ensuring data collection and usage is transparent. It should be able to make occupants feel comfortable and calm, without being invasive or critical about the occupants’ behaviours. 12. The location and quality of stimuli should be clear and designed to meet occupants’ needs based on their role in the vehicle (i.e. driver, front passenger, rear passenger). 13. System stimuli should be designed as per individual needs, ensuring that the experience of one occupant does not cause hindrance to the experience of others. 14. The system should be congruent with a user’s mental model of an in-vehicle experience, as an unexpected response from the system can lead to mistrust and reduced user interaction. 15. The system should provide limited but sufficient information about its status to avoid any confusion and maintain a balance between occupant safety and experience.

Conclusions

The proposed heuristics are intended to support driver wellbeing but can also be applied as a discount method to evaluate in-vehicle AmI, more generally, and may be adapted to other situations in which AmI supports a critical primary task. Further work will seek to refine/validate the heuristics and explore other novel methods to evaluate the impact of in-car AmI and inform design.

References

Augusto, J.C., Callaghan, V., Kameas, A., Cook, D. and Satoh, I., 2013. Intelligent Environments: a manifesto. Human-centric Computing and Information Sciences, 3: 12, 2013. *Springer*. DOI, 10, pp.2192-1962.

Dunne, R., Morris, T. and Harper, S., 2021. A survey of ambient intelligence. *ACM Computing Surveys (CSUR)*, 54(4), pp.1-27.

Fortune, K., 2024. *Wellness on wheels: can a new generation of cars make you healthier?* Available online: https://unfilteredonline.com/wellness-on-wheels-can-a-new-generation-of-cars-make-you-healthier/ (Accessed 21.10.24).

Nieslen, 2024 *10 Usability Heuristics for User Interface Design*. Available online: https://www.nngroup.com/articles/ten-usability-heuristics/ (Accessed: 21.10.24)

Nielsen, J. and Molich, R., 1990, March. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 249-256).

Stephanidis, C., Antona, M. and Ntoa, S., 2021. Human factors in ambient intelligence environments. *Handbook of human factors and ergonomics*, pp.1058-1084.

Do Hands-On or Hands-Off Drivers Perform Better During a Level-2 Silent Failure?

Amaad Hussain, David R. Large, Catherine Harvey

Human Factors Research Group, University of Nottingham, UK

SUMMARY

In a driving context, a silent failure (SF) occurs when the automated vehicle system is not able to operate but does not recognise its reduced capacity. As a result, it does not request human intervention or alert the user to the failure; the human driver must therefore recognise the need for their input and provide it in an appropriate and timely manner. Recent legislation is beginning to relax the stipulation for drivers to keep their 'hands-on' the steering wheel during level 2 (L2) automation. This study therefore explores if and how hand position ('hands-on' or 'hands-off') influences drivers' reaction time and the quality of their intervention following a silent failure at L2 automation. Results show that 'hands-on' drivers were significantly quicker to respond to the SF, with the first steering input occurring in 3.7s, on average, amongst this cohort, compared to 8.5s for 'hands-off' drivers. Significantly, all 'hands-on' drivers provided an active control input before their vehicle drifted across the lane boundary.

KEYWORDS

Level 2 automation, silent failure, hands-on, hands-off.

Introduction

There has been considerable research interest in understanding and improving the transfer of control between vehicle automation and the incumbent driver. Much of this work focusses at level 3 'conditional' automation (or higher) (SAE, 2021). It therefore applies to 'next generation' vehicles, where the driver is able to disengage from all aspects of the driving task during periods of automation and must therefore re-engage with the driving task, prior to taking manual control. In contrast, current generation vehicles are increasingly offering level 2 driver control and assistance systems (L2-DCAS) (SAE, 2021). L2-DCAS requires drivers to supervise the vehicle 'completely and continuously' while the system is in control and are therefore generally considered to be 'hands on, eyes on' systems (i.e. the driver is required to keep their hands on the steering wheel and their eyes on the road *at all times*). In 2023, the UK Vehicle Certification Agency approved the Ford BlueCruise system as the first '*hands off*, eyes on' DCAS (also referred to as L2++) in the UK (PACTS, 2024). Although such functionality is currently limited to specific roads ('blue zones') (Ford, 2022), there is a tacit assumption that 'hands off, eyes-on' is sufficient to keep drivers appraised of the driving situation and in physical readiness to resume control if/when required. Allowing drivers to take their hands off the steering wheel during automation purportedly reduces workload (Naujoks et al., 2015), making this an attractive proposition. However, it also disconnects the driver from the perception-action cycle (Cutsuridis et al., 2011), and, practically, means that their hands are now able to engage in non-driving related tasks (NDRTs) (Damböck et al., 2013), should they so desire; this, in itself can direct their attention away from the road further (Gershon et al., 2023).

It is also often assumed that in any situation outside of its control, an automated vehicle (at any level of automation) will identify its own limitation (or a malfunction) and call the driver to action. However, this will not always be the case. Indeed, it is feasible that the automation could fail in a manner not recognised by the system. It will therefore not issue a take-over request (TOR) and the driver will not be advised of the need to resume control. This is routinely referred to as a 'silent failure' (SF) and could occur if a sensor misses or misinterprets lane markings, for example, resulting in uncontrolled lane drift (see: Lambert, 2018). SFs are largely unpredictable (to both the system and the incumbent driver), and thus drivers' responses are unknown, although are expected to be slower, and the consequences and repercussions more severe, than during a planned or alerted take-over (Mole, et al, 2020).

Method

The aim of the current study is to investigate drivers' behaviour in response to a SF, as determined by their hand position ('hands-on' or 'hands-off') during the preceding L2 automation. The study took place in the University of Nottingham Human Factors driving simulator (Figure 1). AVSimulation SCANeR software was used to create a three-lane UK-motorway scenario with traffic distributed across all lanes. The simulator was modified to provide an authentic L2-DCAS experience lasting approximately 15 mins and culminating in a SF. During automation, the car performed 6 automated lane changes (effectively manoeuvring between all 3 lanes as dictated by the surrounding traffic) and braked/accelerated in response to the behaviour of other vehicles.

Figure 1: Driving simulator with AVSimulation SCANeR software

Participants

Fourteen drivers took part (7M, 7F; ages, 18-34; mean driving experience, 3.9yrs). All participants were specifically instructed to, "supervise the system automation as you would if you were manually driving", although they were not made aware of the upcoming silent failure. Seven participants were asked to keep their hands on the wheel ('hands on, eyes on'), while the remaining seven were permitted to remove their hands from the steering wheel ('hands off, eyes on'); no specific instructions were provided to the latter, 'hands-off' participants regarding whether they could engage in non-driving related tasks, for example. However, these drivers were reminded to return their eyes to the road if their off-road glances extended beyond approximately 2s, although this was not enforced precisely – the aim was to re-direct drivers' visual attention to the road if they appeared to be becoming 'inappropriately distracted' (e.g. using their mobile phone), as a driver monitoring system might behave.

In-Vehicle HMI

Prior to the SF, the system provided lateral and longitudinal control and informed the driver of its behaviour via an in-vehicle HMI, for example, declaring ‘vehicle performing manoeuvre’ using text and iconography to accompany a lane change (Figure 2). The HMI also reminded ‘hands-on’ drivers, via a text message, to replace their hands on the steering wheel if they inadvertently removed them during automation and to remind drivers in both groups to return their eyes to the road if they became distracted (Figure 2). The messages on the HMI were preprepared and controlled manually, as the situation dictated, by the experimenter from the control room.

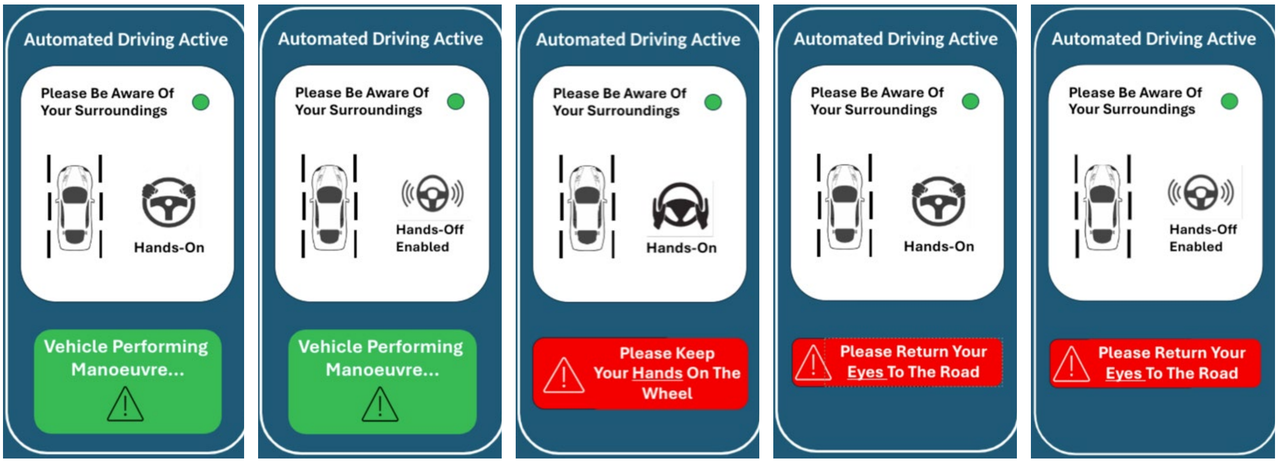


Figure 2: Example HMI screens used during periods of automation, showing (left to right): vehicle performing a manoeuvre (i.e. a lane change) for ‘hands-on’ and ‘hands-off’ drivers, a reminder for ‘hands-on’ drivers to return their hands to the steering wheel, and reminders to ‘hands-on’ and ‘hands-off’ drivers to return their eyes to the road

Silent Failure

The journey lasted approximately 15 minutes, with the SF occurring approximately 20s after the final announced manoeuvre. The SF manifested as a loss of both lateral (steering) and longitudinal (speed) control. It occurred when the vehicle was in lane 3 (right lane) travelling at approximately 70mph and had no accompanying notification on the HMI. Thus, drivers were required to detect the failure and take whatever corrective action they deemed appropriate. They were obviously not aware that a SF would occur, and no instructions had been provided regarding what to do if such a situation were to occur. If no driver intervention took place, the vehicle drifted to the left across lane 2 and into lane 1, and came to a stop in lane 1 after approx. 40s. Nearby, a ‘break-down’ truck was attending to a stationary vehicle on the hard shoulder. This was in the vehicle’s initial trajectory (i.e. when control was surrendered), and intended to present an apparent collision hazard, although no collision would have actually taken place even with no driver intervention (Figure 3). In addition, all lanes remained active with other vehicles present (as they had been throughout the journey), and thus, these vehicles also presented potential collision hazards. However, nearby vehicles acted intelligently around the ego vehicle (e.g. braking to avoid colliding with it).

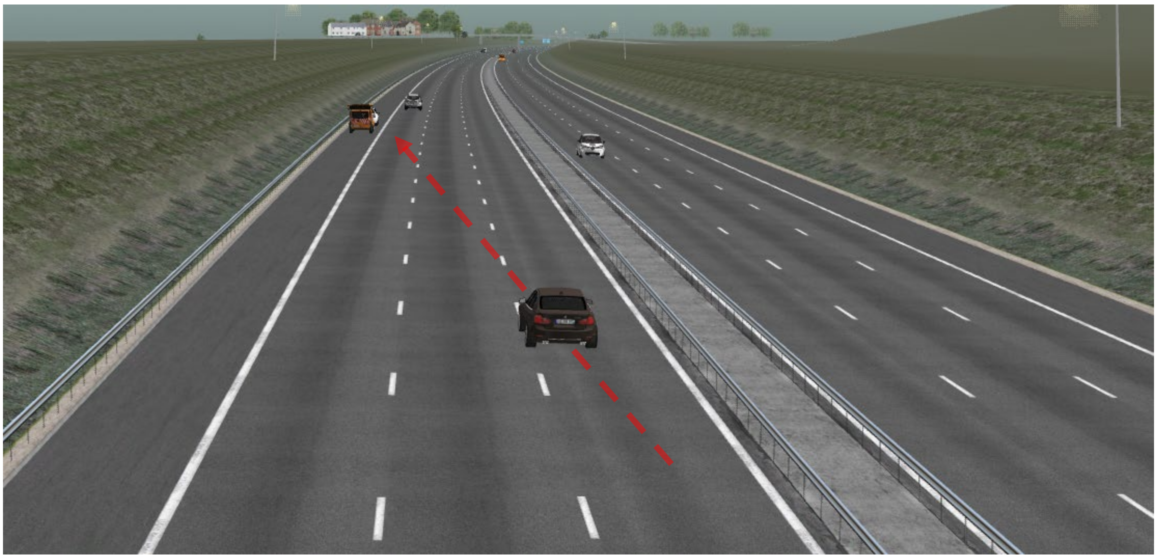


Figure 3: Image showing position of ego vehicle (black BMW) shortly after silent failure, with break-down truck attending to vehicle on hard shoulder in the ego vehicle's initial trajectory

Measures

Videographic, simulator and eye-tracking data were captured to determine participants' first response, and their vehicle control and visual behaviour following the SF. A post-study questionnaire with open-ended questions captured insights into how participants perceived their level of success at intervening and how well they felt they were able to maintain their attention on the road during automation. Unfortunately, two participants' simulator data were corrupted and thus, these participants' data were removed prior to analysis (one from each group).

Results and Discussion

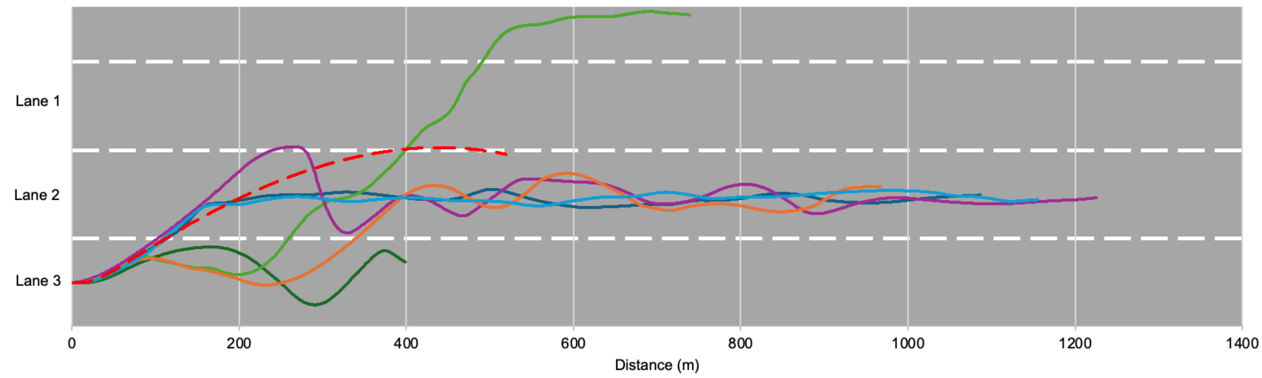
Drivers' response and their reaction time was determined by their first active input to any primary control (steering, accelerator or brake), using data obtained from the simulator. For all participants (i.e. in both groups), the first active input was to steering. This manifested as either a counterwise corrective manoeuvre (presumably if they were aiming to keep the vehicle in the current lane) or a positive reinforcement of the current trajectory (if accepting the move into the adjacent lane and aiming to position their vehicle appropriately within that lane). Notably, 'Hands-on' drivers were significantly quicker to respond, with steering input occurring in 3.7s, on average, compared to 8.5s for 'hands-off' drivers (t-test, $p = .02$).

Vehicle trajectories during 60 seconds following the SF is shown in Figure 4. Although perhaps not immediately apparent from the graph, the initial response from all 'hands-on' drivers (Figure 4-top) occurred while their vehicle was still in lane 3 (i.e. had not crossed the lane boundary). Three of these drivers choose to remain in lane 3 by applying a counterwise corrective manoeuvre (although one of these moved to the hard shoulder to terminate their journey shortly afterwards), whereas the remaining 3 'hands-on' drivers actively steered their vehicle into lane 2, where they remained.

In contrast, five out of the six 'hands-off' drivers (Figure 4-bottom), had already crossed the lane boundary and drifted into lane 2 before reacting and providing a control input (one of these drivers subsequently moved back into lane 3 after 20s, or so). It is also notable from Figure 4 that after 30s, or so, the trajectories of 'hands-on' drivers who continued were all clustered neatly in the centre of lane 2, suggesting good lateral control. In contrast, the behaviour of 'hands-off' drivers after the same time period, appears to be more erratic, suggesting poor lateral stability, or wavering in the lane.

The first input to the accelerator or brake pedal occurred later than steering input (27s and 18s post-SF, on average, for 'hands-on'/'hands-off', respectively; as a reminder, steering input had occurred after 3.7s and 8.5s, respectively). This suggests that all drivers perceived (and 'corrected') lateral drift before noticing and responding to the retardation in speed and/or prioritised re-establishing lateral position before resuming longitudinal control. Vehicle speed during the 60s following the SF varied between individuals, with no significant difference, on average, between groups. However, it is notable that there was a tendency for drivers to accelerate above the speed limit (70mph) after taking control, with the maximum speed recorded during this time period of 82mph for 'hands-on' drivers and 90mph for 'hands-off' drivers. Respective mean speeds during the same period were 63mph and 66mph, respectively, although it is noted that some drivers were beginning to decelerate or brake towards the end of this period.

After drivers had accepted control of their vehicle, some chose to continue their journey (as evidenced by initial accelerator input), whereas others evidently planned to slow/stop the car (i.e. initially applied brakes). Three 'hands-on' and four 'hands-off' drivers chose the former strategy, whereas the remaining three/two (hands-on/hands-off, respectively) applied the brake first or allowed the vehicle to slow down passively while still controlling its lateral behaviour (one of these 'hands-on' drivers ultimately steered the vehicle onto the hard shoulder) (Figure 4). It is noted, however, that no specific instructions were provided regarding what to do if/when they were required to resume manual control during this study, and this may have influenced their behaviour.



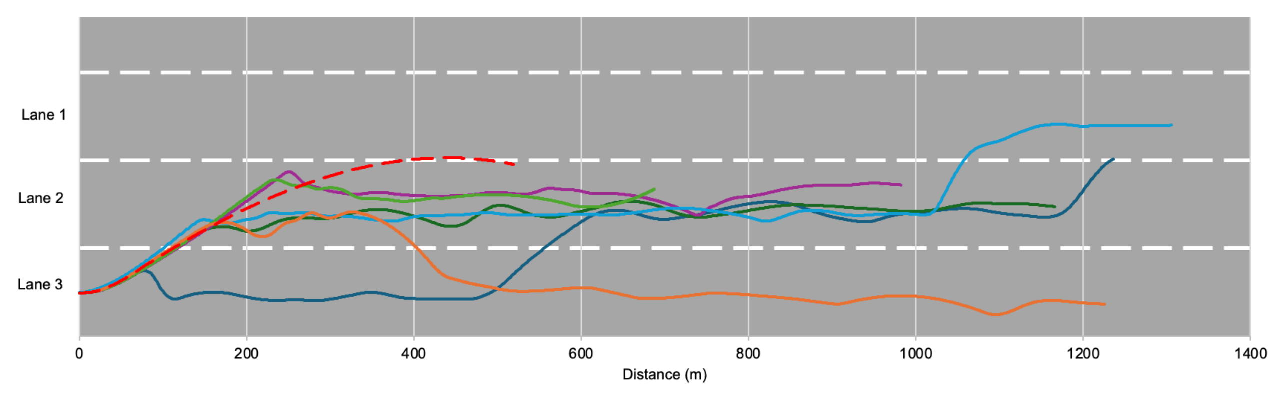


Figure 4: Vehicle trajectories during 60s following silent failure (at distance=0), showing 'hands-on' drivers (top) and 'hands-off' drivers (bottom). Note: dashed red line shows predicted trajectory of vehicle if no driver intervention took place

Visual behaviour was coded using semantic gaze mapping for 2 minutes immediately prior to the SF and 40s post SF. Results show some variability between groups (and indeed, between drivers) during both time periods, but no significant differences, on average, between groups for number and duration of fixations to different areas of interest (road scene, instrument cluster, mirrors etc.). This suggests that visual scanning behaviour had been unaffected, overall, by enforcing hand position.

Responses from the post-study questionnaire suggest that most 'hands-on' drivers believed they reacted 'effectively' when required to resume control: "I noticed the car drifting a bit, so I took control". In contrast, responses from some 'hands-off' drivers suggest that they had become complacent: "I was trusting the car too much", with some recognising that they subsequently 'panicked' when they realised that the automation system had failed and that this impacted on the quality of their takeover: "although I managed to regain control, I did panic slightly which resulted in oversteering initially and accelerating above the speed limit." In their comments, 'hands-on' drivers also generally expressed a more active role in supervising the car than 'hands-off' drivers. For example, one 'hands-on' driver commented that they applied, "the same level of visual attention as I would in fully manual driving", whereas 'hands-off' drivers' responses suggested they were more passive: "I felt my attention to the road was reduced. I felt distracted when [the car] said it was doing a manoeuvre."

Finally, a notable observation from videographic data was that several 'hands-off' participants placed their hands on the wheel during automation, ostensibly at times they perceived higher risk, for example, during high traffic flow, lane changes etc. This occurred on 16 occasions (or 2.7 times per participant), despite no specific instruction to this effect, and suggests that 'hands-on' was perceived as 'safer' by these drivers. Additionally, several drivers were observed moving their foot over the brake or accelerator pedal during automation, presumably in preparation for intervening. This occurred on 4 occasions for 'hands-on' drivers and once for 'hands-off' drivers.

Discussion

There is a tacit assumption that an automated vehicle (AV) will alert the driver and issue a take-over request (TOR) if it encounters a situation that is outside of its control. However, it is possible that the AV will not be aware of its reduced capacity (for example, due to the loss of lane demarcations or a sensor malfunction) and will therefore be unable to call the driver to action – a so-called silent failure (SF). Meanwhile, the vehicle is effectively 'unattended', with no active driver or automation system in control. This is extremely problematic, not least if the vehicle was travelling at 70mph in lane 3 of a motorway when the SF occurred. Understanding how drivers will naturally respond in this situation is therefore important in order to guide the design and development of technological solutions as well as to create supporting guidance and appropriate regulations etc.

In addition, given the general drive towards L2++ systems, and the associated relaxation of the current stipulation for drivers to keep their 'hands-on' the steering wheel during periods of automation, it remains to be seen if drivers are able to respond as effectively (in terms of their reaction time and quality of intervention) with their 'hands-off' compared to drivers with their 'hands-on' the steering wheel during periods of L2++ automation. While the study was 'between-subjects' (out of necessity, as it was exploring drivers' first response to an unexpected situation), and the number of participants was limited, it nevertheless provides some useful preliminary data to highlight this predicament.

Indeed, results show that 'hands-on' drivers were significantly quicker to detect and react to a SF, taking active (lateral) control of their vehicle in effect *before* it moved out of its current lane. In comparison, 'hands-off' drivers had typically drifted into lane 2 before responding and intervening. It is also notable that in both situations, drivers first action was to correct lateral deviation, as opposed to applying longitudinal (i.e. speed) control. Moreover, 'hands-off' drivers' vehicle control remained erratic even after 60s post SF.

Conclusion

Results from the study suggest that by keeping their hands on the steering wheel during L2 automation, drivers are better attuned to changes in the behaviour of their vehicle during automation than drivers who are permitted to remove their hands from the steering wheel. Consequently, the former, 'hands-on' drivers are better prepared to resume control if/as demanded by the vehicle or road situation, most notably, in situations such as a silent failure, where the need to take over control is both unexpected and unannounced. Future work should continue investigations with a larger cohort of participants and in a variety of driving scenarios.

References

Cutsuridis, V., Hussain, A. and Taylor, J.G. eds., 2011. Perception-action cycle: Models, architectures, and hardware. Springer Science & Business Media.

Damböck, D., Weißgerber, T., Kienle, M. and Bengler, K., 2013, October. Requirements for cooperative vehicle guidance. In 16th international IEEE conference on intelligent transportation systems (ITSC 2013) (pp. 1656-1661). IEEE.

Ford. 2022. Ford BlueCruise: Hands Free Driving Technology | Ford UK. Available at: https://www.ford.co.uk/technology/driving-assistance/ford-bluecruise

Gershon, P., Mehler, B. and Reimer, B., 2023. Driver response and recovery following automation initiated disengagement in real-world hands-free driving. Traffic injury prevention, 24(4), pp.356-361.

Lambert, F., 2018. Tesla Autopilot confuses markings toward barrier in recreation of fatal Model X crash at exact same location. URL: https://electrek.co/2018/04/03/tesla-autopilot-crash-barrier-markings-fatal-model-x-accident/.

Mole, C., Pekkanen, J., Sheppard, W., Louw, T., Romano, R., Merat, N., Markkula, G. and Wilkie, R., 2020. Predicting takeover response to silent automated vehicle failures. PLoS One, 15(11), p.e0242825

Naujoks, F., Purucker, C., Neukum, A., Wolter, S. and Steiger, R., 2015. Controllability of partially automated driving functions–does it matter whether drivers are allowed to take their hands off the steering wheel? Transportation research part F: traffic psychology and behaviour, 35, pp.185-198.

PACTS. 2024. Hands-off driving assistance systems steam ahead - The Parliamentary Advisory Council for Transport Safety (PACTS). Available at: https://www.pacts.org.uk/hands-off-driving-assistance-systems-steam-ahead/

SAE International. 2021. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles J3016_202104. Available at: https://www.sae.org/standards/content/j3016_202104/

Enhancing Inclusivity and Safety in Self-Driving Taxis

Clare Mutzenich[1], Emily Stobbs[1], Ahmed Ehab Abdelsalam[2] & Gary Burnett[2]

[1]Lacuna Agency, [2]Loughborough University

SUMMARY

This study explored the challenges individuals with protected characteristics face during emergencies in self-driving taxis and how inclusive design can enhance safety. Automated Passenger Services (APS) offer significant mobility benefits, but the absence of a driver shifts safety responsibilities to passengers and automated systems. Key emergency tasks—such as contacting emergency services, making evacuation decisions, and interacting with transport operators—vary in complexity depending on passengers' physical, cognitive, and sensory abilities. A qualitative research approach, including focus groups and VR simulations, was used to understand the actions passengers must take during emergencies and the system features required to support them. The study identified that users with disabilities, older adults, and younger passengers face significant barriers in these situations, while personal safety concerns related to gender, race, religion, or sexual orientation influence decision-making. Findings indicated that APS must integrate adaptive human-machine interfaces (HMIs), real-time operator support, and tailored accessibility features to ensure safety for diverse users. Recommendations include manual door overrides, multimodal alerts, and AI-driven emergency assistance. As the APS sector evolves under the Automated Vehicle Act 2024, collaboration with industry stakeholders will be essential. Engaging user groups, transport operators, and technology providers will help integrate inclusivity into self-driving taxi services, improving safety and accessibility while setting a new standard for user-centred innovation.

KEYWORDS

Inclusivity, Emergency Design, Self-Driving Vehicles, Virtual Reality, Protected Characteristics

Introduction

The increasing development and deployment of Automated Passenger Services (APS), such as self-driving taxis, presents significant opportunities for enhancing mobility, accessibility, and safety. In April 2023, the Centre for Connected and Autonomous Vehicles (CCAV) identified a lack of established guidance relating to how different user groups interact with self-driving transport systems across a variety of tasks, such as booking, payment, and wayfinding. As the deployment of APS expands, it is essential to consider not just how users interact with the system under normal conditions, but also how they will navigate unforeseen situations, particularly in emergencies. Emergency situations within APS present unique challenges, particularly due to the absence of a driver, which places greater responsibility on users. Key user tasks during emergencies may include contacting emergency services, communicating with the transport operator, and making evacuation decisions. Emergencies, such as vehicle collisions, floods, or medical crises, present significant challenges, particularly for vulnerable populations. The complexity of these tasks may vary based on individual passenger needs, capabilities, and age. Historically, transport systems have overlooked diverse user needs—for example, crash test dummies were primarily modelled on male bodies, neglecting women, children, and people with disabilities (Schiebinger, 2014).

The rise of self-driving taxis heralds a new era in personal mobility, promising increased accessibility and convenience (Department for Transport, 2019). However, these benefits are contingent on their ability to accommodate all users, including those with protected characteristics under the Equality Act 2010 (Equality Act, 2010). The Act mandates that public services, including transport, provide equitable access and support to individuals with diverse needs. The Act identifies nine specific characteristics that are legally protected from discrimination, harassment, and victimisation, referred to in this report as protected characteristics. These are:

- **Age:** Covers people of all age groups, ensuring protection against age-related discrimination.
- **Disability:** Includes physical and mental impairments that have a substantial and long-term impact on an individual's ability to carry out daily activities.
- **Gender Reassignment:** Protects those who are undergoing, have undergone, or are considering a process to reassign their gender.
- **Marriage and Civil Partnership:** Protects against discrimination based on marital status or civil partnership, but only in employment contexts.
- **Pregnancy and Maternity:** Protects individuals from discrimination during pregnancy and up to 26 weeks after giving birth, including breastfeeding.
- **Race:** Encompasses colour, nationality, and ethnic or national origins.
- **Religion or Belief:** Protects people with religious beliefs, non-religious beliefs (e.g., atheism), or philosophical beliefs (e.g., environmentalism), provided they are worthy of respect in a democratic society.
- **Sex:** Protects individuals from discrimination based on being male or female.
- **Sexual Orientation:** Covers heterosexual, homosexual, and bisexual orientations.

In traditional taxis, human drivers may address specific requirements during emergencies, such as assisting passengers with mobility impairments, ensuring safe evacuation for wheelchair users, or supporting pregnant passengers. For older adults, drivers may provide reassurance or additional physical assistance, and for passengers with sensory impairments, clear verbal communication is often essential. These practices align with the Equality Act by ensuring that individuals with protected characteristics are not disadvantaged in emergencies.

Research has highlighted several key challenges faced by individuals with protected characteristics in transport systems. Individuals with mobility and cognitive impairments often struggle with tasks like recognising hazards, securing seatbelts, and stowing luggage, requiring additional assistance (Schräder, 2019; Millonig & Fröhlich, 2018). Anxiety, lack of confidence, and mental health conditions (e.g., depression, agoraphobia) can limit travel frequency and independence, particularly for those with cognitive and behavioural impairments (Mackett, 2017). Visually impaired passengers may face difficulties navigating vehicles and rely heavily on digital tools for journey planning. In automated systems, traditional support tasks, like boarding and exiting, would need to be automated to ensure accessibility (Low, 2020; Millonig & Fröhlich, 2018).

This project seeks to fill knowledge gaps through a human factors and ergonomics approach to investigate user tasks and vehicle interactions during emergency scenarios and their associated requirements. Lacuna Agency worked in partnership with Loughborough University, combining expertise in Human Factors research and VR simulations. This collaboration, together with DfT and CCAV aimed to provide evidence-based insights to inform emergency accessibility protocols. The research was commissioned to explore the needs of diverse populations using APS, focusing on inclusivity and accessibility, as well as the requirements for managing tasks during emergency situations. Particular attention was given to the experiences of users with protected characteristics as defined under the Equality Act 2010. This study leverages VR technology to simulate emergencies and gather insights into user behaviours, perceptions, and needs. By focusing on inclusivity and safety, the research aligns with the conference theme of enhancing human performance through the integration of technology and human-centred design.

Methodology

This study used a qualitative research approach, incorporating focus groups and VR simulated self-driving taxi experiences to investigate user needs and behaviours during emergency scenarios. A quota sampling strategy was employed to ensure diverse participant demographics, with particular attention to individuals falling under the protected characteristics defined by the Equality Act 2010. To ensure that systems effectively support user needs, task analysis was applied to identify the cognitive and physical actions participants reported they would need to perform in these emergencies.

A total of 91 participants were recruited for this study, including 10 children (aged 8-17yrs old) who participated with their parents. The sample were recruited from cities and towns located in the Midlands region of England, UK (Leicester, Nottingham, Loughborough, and Birmingham), reflecting both urban and rural areas. To qualify, all participants were required to be able to use public transport independently and 80% of participants needed to use a taxi service more than twice a month, with more flexibility allowed for accessibility groups.

Participants were recruited to reflect a broad range of protected characteristics under the Equality Act 2010, ensuring representation across various age groups, genders, ethnicities, and social grades. Recruitment methods prioritised intersectionality, aiming for participants with two or more protected characteristics where possible, to provide nuanced insights into diverse user needs. The protected characteristic of disability was split into categories to represent distinct user groups to understand specific accessibility needs in this group: physical impairment, hearing impairment, vision impairment, and neurodivergent (Autism Spectrum Disorder - ASD, Dyslexia, Dyspraxia, Attention deficit disorder - ADD). The study was conducted in adherence to the ethical standards set by Loughborough University Ethics Committee to ensure participant safety, inclusivity, and data privacy.

Participants could participate in either “solo” sessions in which they participated alone, or “social” sessions in which they participated alongside three others (total of four participants). Participants under the ages of 18 years old were only permitted to take part in social sessions with their parents for safeguarding reasons. The inclusion of both solo and social VR sessions in this study was designed to reflect the varied contexts in which passengers might experience emergencies in self-driving taxis. These scenarios aimed to capture the diverse psychological, behavioural, and emotional responses that could arise depending on whether a passenger is alone or accompanied.

The emergency scenarios for the VR simulations were developed in collaboration with CCAV through a workshop involving key stakeholders from transport charities and organisations. The workshop helped to prioritise the types of emergencies to focus on, selecting those deemed the most common in transport situations, likely to occur, or of significant concern to potential future users of APS. This research prioritised the inclusion of voices from groups often underrepresented in transport research, with a focus on ensuring there were no barriers for people with disabilities or other protected characteristics to participate. Several measures were implemented throughout the study to support participants. For those with sensory issues or neurodivergence, such as ADD, regular breaks were provided to reduce fatigue and prevent cognitive overload. A designated “chill-out” area was set up for participants to rest if needed. Sensory adjustments included dimming the lighting and reducing the number of researchers in the room, allowing participants to engage more comfortably. These steps ensured participants were able to engage without feeling overwhelmed.

Carers and support staff were encouraged to accompany participants, particularly those with mobility impairments or cognitive disabilities, to ensure they could fully engage with the study. Carers could either stay in the room to support participants directly or remain in the designated chill-out area for additional assistance. The venue was made fully accessible, with lift access and adjustments to the physical setup, allowing participants in wheelchairs to remain seated throughout the study without needing to transfer. To address transportation challenges, detailed instructions, including maps and parking information, and a dedicated host were provided in advance to ensure participants with mobility impairments or cognitive challenges could navigate the venue easily. For pregnant participants or those with young children, flexible scheduling allowed for more frequent breaks and timing adjustments to meet their needs. These efforts along with others were put in place to ensure that participants, regardless of their background or needs, could fully engage in the study without facing unnecessary barriers. This approach reflects the research team's commitment to inclusivity and accessibility, ensuring that all users could participate meaningfully.

Emergency scenarios

The study focused on six carefully crafted emergency scenarios, each representing a distinct type of emergency, including internal incidents (e.g., medical emergencies), external threats (e.g., vehicle collisions), environmental hazards (e.g., flooding), and interpersonal situations (e.g., passenger altercations). The six scenarios were:

- **Pedestrian interaction:** A pedestrian attempts to open the taxi door while the vehicle is stopped at a red light.
- **Medical emergency:** Participants share the self-driving taxi with an unfamiliar passenger (a VR avatar) who becomes unwell during the journey.
- **Incorrect stopping point:** Participants are travelling to the library but the self-driving taxi misses the designated stop and attempts to drop them off further away.
- **Road closure due to flooding:** The taxi is caught in bad weather, where heavy flooding and a barricade force the vehicle to stop.
- **Fire or smoke emergency with door malfunction:** The vehicle catches fire while the participants are on their way home, producing smoke and flames, while the doors fail to open despite the vehicle stopping.
- **Vehicle collision:** The self-driving taxi is hit from behind by another vehicle while stopped at a red light.

Participants experienced three out of the six scenarios, selected randomly, to provide diverse exposure while avoiding cognitive or physical strain from time spent wearing the VR headset. By immersing participants in these VR simulations, the study gathered detailed qualitative feedback on the necessary user actions and cognitive processes, the system features needed to support users, and, if relevant, the protected characteristic considerations related to that step. This innovative approach generated rich qualitative data, offering valuable insights into how to design inclusive and effective emergency protocols for self-driving taxis.

Results

Each of the six emergency scenarios was analysed through task analysis, examining the actions, thoughts, and system interactions required by passengers in self-driving taxis. The study assessed whether passengers could manage these tasks independently or needed additional support. Insights from interviews and focus groups highlighted key emotional, cognitive, and behavioural challenges. The analysis also explored how protected characteristics, such as age, disability, and cognitive impairments, affected passengers' ability to handle emergencies without a driver. Some issues were universal, while others were specific to particular emergency scenarios. The intersection of multiple protected characteristics was also examined to identify unique barriers.

A master task framework was developed to outline the key emergency tasks required of passengers, given the absence of a driver, shifting responsibility to automated systems and the users.

1. **Awareness and recognition** – Passengers must detect and understand potential risks. Unlike traditional taxis, where drivers provide verbal warnings and alerts, self-driving taxis rely on system notifications (audio, visual, or sensory cues). These alerts must be clear and accessible to all passengers, particularly those with sensory or cognitive impairments.
2. **Assessing the situation** – Passengers must evaluate the severity of the emergency and potential risks. In traditional taxis, drivers assess the situation and take appropriate action. In self-driving taxis, passengers rely on system-provided information to determine their next steps, such as recognising hazards like fire, flooding, or an unwell co-passenger.
3. **Decision making** – Without a driver to make emergency decisions, passengers must act independently or with system guidance. Features such as emergency stop buttons, door overrides, and clear exit instructions can support users. Ensuring that these systems are intuitive and accessible is essential for a diverse range of passengers.
4. **System interaction** – Passengers must engage with system controls to execute their emergency response. Traditional taxis require little direct system interaction, as drivers manage vehicle functions. In self-driving taxis, passengers may need to activate emergency stop mechanisms, issue voice commands, or override door locks. Interfaces must be user-friendly and adaptable for individuals with different abilities.
5. **Communication** – Passengers must contact external parties, such as operators, emergency services, or other passengers. Traditional taxis rely on drivers to handle communication. In self-driving taxis, passengers will need robust communication tools, such as intercoms, pre-recorded emergency messages, or live operator support. Multilingual and accessibility features will enhance inclusivity.
6. **Post-incident actions** – Passengers must manage the aftermath of an emergency, such as arranging alternative transport or seeking further assistance. Traditional taxi drivers help facilitate these steps. In self-driving taxis, automated assistance, live operator support, and accessible travel updates will be crucial for ensuring passengers feel supported post-incident. These findings underscore the importance of inclusive safety features and tailored emergency protocols that address the diverse needs of passengers, ensuring self-driving taxis are safe and equitable for all users.

Key passenger needs during emergencies in self-driving taxis include clear and accessible alerts, such as audio-visual notifications, to ensure all users can recognise potential dangers. Safe and accessible exit options, including emergency stop buttons, door overrides, and ramps, are essential for a quick and efficient evacuation. Passengers also require guidance on how to interact with external parties, such as emergency services or operators, through intercoms and emergency contact systems. Emotional reassurance is another critical factor, with structured system messages or live operator support helping to reduce anxiety and improve decision-making under stress.

To facilitate effective incident reporting and help-seeking, systems should include voice assistance and subtitles, ensuring accessibility for all users. Real-time alternative travel arrangements and accessible navigation tools will support passengers after an incident, helping them continue their journey safely. Proactive hazard detection, such as fire, flooding, and collision sensors, can further enhance passenger safety by enabling early intervention. Additionally, emergency protocols must be culturally and religiously sensitive, incorporating multilingual support and gender-sensitive communication to accommodate diverse user needs. These findings underscore the importance of inclusive emergency protocols in self-driving taxis. Accessibility considerations are particularly crucial for passengers with sensory, physical, or cognitive challenges, ensuring that all users can effectively respond to emergencies without the presence of a driver.

Discussion

This research contributes to the broader discourse on enhancing human performance by addressing inclusivity in autonomous transport design. By focusing on protected characteristics, it ensures that self-driving taxis accommodate diverse user needs during emergencies. The findings highlight the intersectionality of characteristics, such as disability and gender, in shaping perceptions of safety and usability. The findings offer crucial insights into the challenges passengers face during emergencies in self-driving taxis, particularly those from groups more likely to experience vulnerability or exclusion. Without a driver to provide guidance, mediate interactions, or offer physical support, passengers must rely entirely on automated systems and interfaces. This shift introduces unique difficulties, particularly for individuals with disabilities, older adults, and younger passengers, who may encounter barriers due to physical, cognitive, or sensory limitations. Beyond practical challenges, personal safety concerns—shaped by factors such as race, religious beliefs, gender, or sexual orientation—also influence how individuals perceive and respond to emergencies. Barriers experienced by different user groups included:

Disability: The absence of a driver removes an essential source of immediate assistance for passengers with disabilities. Physical impairments make features such as automated door unlocking, reliable ramps, and clear evacuation paths critical for ensuring accessibility without human intervention. Vision-impaired users, who might typically rely on a driver for situational context, require voice-guided navigation and audible alerts. Hearing-impaired passengers, unable to engage in verbal communication, depend on visual aids and real-time updates. Neurodivergent users, who may benefit from the calming presence of a driver, require structured and intuitive systems to minimise overstimulation and distress.

Sex: The lack of a driver increases feelings of vulnerability, particularly among female and non-binary passengers. Automated safety features such as panic buttons, alternative exits, and live operator support could help replace the protective presence of a human driver. For male passengers, the absence of a driver who might otherwise mediate interactions raised concerns about accountability. Transparent monitoring systems, including CCTV and audio recording, were seen as necessary to ensure fair treatment and prevent misinterpretation during emergencies.

Sexual orientation: LGBTQIA+ passengers expressed concern about the absence of a driver to deter potential harassment or discrimination. High-quality CCTV and live operator support were identified as essential for ensuring safety in high-risk situations. Automated systems that detect and mitigate discriminatory behaviour could provide reassurance, replacing the role a driver might play in preventing or intervening in incidents of bias.

Religion: For passengers from religious backgrounds, visible safety features such as CCTV were regarded as important deterrents against prejudice. The lack of a driver to offer situational clarity in emergencies underscored the need for alternative exits and culturally sensitive communication systems to accommodate diverse needs.

Age: Younger and older passengers were particularly affected by the absence of a driver. Younger passengers (8–17 years), who might typically look to a driver for reassurance or authority, require structured guidance and live operator support. Older passengers (65+ years), who may rely on drivers for physical assistance, need accessible features such as low-threshold doors, wide ramps, and tailored medical alerts. Middle-aged passengers benefit from clear, step-by-step system guidance to navigate emergencies effectively.

Pregnancy and maternity: Pregnant passengers highlighted the absence of a driver as a key factor in their sense of vulnerability during emergencies. Features such as automated door unlocking, low-threshold doors, and ramps were seen as essential for safe evacuation without physical assistance. Live operator connections could replicate the reassurance a driver might provide, while integrated medical response systems could replace the driver's role in assisting with health concerns.

Race One participant expressed heightened anxiety about the risk of misjudgement or prejudice in the absence of a driver to mediate interactions. Real-time video and audio monitoring were considered crucial for ensuring fairness and accountability. Automated alerts and impartial instructions could help reduce bias in high-pressure scenarios.

Gender reassignment: For transgender passengers, the absence of a driver removed an important layer of security. Real-time video monitoring and live operator support could offer reassurance and accountability. The use of pre-filled diversity profiles might allow emergency responders to access sensitive information in a way that ensures appropriate and respectful assistance.

Marriage and civil partnership: This characteristic did not significantly influence user needs, as the absence of a driver did not intersect meaningfully with participants' experiences.

Digital exclusion: Passengers with limited digital literacy faced additional challenges due to the reliance on automated systems. Manual buttons, physical intercoms, and simple navigation tools could provide essential support for these users. Live operator connections and non-digital options are key to ensuring effective access to emergency assistance.

Limitations of the study

While the study provided valuable insights into how protected characteristics shape passengers' experiences in self-driving taxis, several limitations should be acknowledged. One challenge was the reliance on individual perspectives to represent broader demographic groups. For instance, the study included only one transgender participant, making it difficult to generalise findings across the wider transgender community. Future research should seek a more diverse participant base to capture a broader range of experiences, potentially by offering home-based sessions to improve accessibility.

Participants were also limited to experiencing only three of the six VR scenarios to minimise fatigue. While this approach helped maintain engagement, it may have restricted the relevance of the scenarios to certain participants' protected characteristics. Future studies could explore longer or staggered sessions to ensure all relevant challenges are captured. Additionally, participants often made assumptions about scenario elements, which may have influenced their responses. For example, in the pedestrian interaction scenario, many assumed the VR pedestrian was male, despite it being designed as gender-neutral. This highlights the importance of designing avatars with explicitly diverse identities in future studies.

Another key finding was that participants sometimes spoke on behalf of others rather than reflecting on their own experiences. Many described how an older or pregnant person might feel in a situation, rather than focusing on how their own protected characteristics influenced their response. While this approach highlights a positive consideration for inclusivity, it also obscured some of the unique challenges individuals faced. Future research could provide clearer instructions to encourage self-reflection. A final challenge was the overrepresentation of older participants, who were often more vocal in their feedback. While their insights were valuable, they may have overshadowed the experiences of younger participants. Ensuring balanced representation across age groups in future studies could help address this.

Recommendations for future self-driving taxi design

The absence of a human driver places new responsibilities on both passengers and automated systems, particularly in emergency situations where passengers must now perform tasks traditionally handled by a driver. The findings highlight that these challenges are both practical—relating to physical, sensory, or cognitive barriers—and psychological, influenced by concerns around safety, trust, and bias. To address these challenges, future automated passenger service (APS) designs should prioritise the following:

- **Accessible vehicle design**: Features such as manual door overrides, weather-resistant ramps, and ergonomic seating would help ensure all passengers can evacuate safely in an emergency.
- **Multimodal communication systems**: Voice, text, and visual instructions should be used to accommodate sensory and cognitive differences, with universally recognisable icons to aid decision-making under stress.
- **Personalised safety features**: Passengers should be able to pre-fill accessibility and safety preferences during booking, allowing the system to provide tailored guidance in emergencies.
- **Proactive emergency planning**: Embedded medical tools, AI-powered evacuation plans, and live operator connections could compensate for the lack of human intervention.
- **Enhanced security and monitoring**: Real-time video and audio monitoring, along with automated alerts, would provide reassurance to passengers concerned about personal safety.
- **Cultural and religious sensitivity**: Systems should be designed to accommodate diverse backgrounds, offering multilingual support and culturally appropriate communication methods.

Conclusion

This study highlights the unique challenges individuals with protected characteristics—such as disabilities, gender, age, and cultural or religious identities—face during emergencies in self-driving taxis. These challenges present opportunities to innovate and collaborate with industry to ensure APS are safe, inclusive, and adaptable to diverse user needs. The study aimed to identify emergency tasks passengers may need to perform, determine when support is most needed, and propose inclusive design considerations. Findings suggest that adaptive HMIs, robust communication systems, and proactive emergency planning can enhance safety and trust in APS. Without a driver, the responsibility for safety shifts to passengers and automated systems, requiring user-centred solutions.

This shift creates an opportunity to rethink how support is provided. Automated systems must take on a more proactive role, ensuring equitable responses to physical, cognitive, and emotional challenges. As the APS sector evolves under the Automated Vehicle Act 2024, collaboration with industry stakeholders will be essential. Engaging user groups, transport operators, and technology providers will help integrate inclusivity into self-driving taxi services, improving safety and accessibility while setting a new standard for user-centred innovation. Using VR simulations, this study explored emergency scenarios in a controlled yet realistic setting, identifying safety and accessibility needs before widespread deployment. By incorporating voices often overlooked in transport planning—such as those of disabled, neurodivergent, and faith-based users—the study underscores the importance of inclusive design. These insights provide a foundation for refining APS development, ensuring self-driving taxis are equitable, trusted, and responsive to the needs of all passengers.

References

Department for Transport. (2019). The Inclusive Transport Strategy: Achieving Equal Access for Disabled People. Retrieved from https://www.gov.uk/government/publications/inclusive-transport-strategy/the-inclusive-transport-strategy-achieving-equal-access-for-disabled-people

Equality Act. (2010). Available at: https://www.legislation.gov.uk/ukpga/2010/15/contents

Low, R. (2020). Barriers and enablers for visually impaired passengers in transportation systems. Disability and Society Journal, 35(4), 510–525.

Mackett, R. (2017). Travel challenges faced by individuals with cognitive and behavioural impairments. Journal of Transport Accessibility, 29(1), 15–32.

Millonig, A., & Fröhlich, P. (2018). Addressing accessibility challenges in autonomous transport systems. Transportation Research Part A, 113, 20–33.

Schräder, F. (2019). The role of dependents in autonomous vehicle interactions. Journal of Transportation and Dependents, 10(4), 250–270.

Schiebinger, L. (2014). *Gendered Innovations: How Gender Analysis Contributes to Research*. Stanford University Press.

Highlighting Barriers to the Use of Voice Systems in Cars: An Interview Study

Sparsh Khandeparker[1], David R Large[1], Catherine Harvey[1], Gary Burnett[2], Karl Proctor[3], & Chrisminder Hare[3]

[1]Human Factors Research Group, University of Nottingham, UK, [2]Loughborough University, [3]Jaguar Land Rover Research

SUMMARY

Advancements in technology have driven significant growth in voice systems, enabling hands-free interaction and offering notable benefits in the automotive domain by reducing visual-manual distractions. Despite the widespread adoption of voice systems in home settings, their use in driving contexts remains limited. This study aims to understand the barriers to using voice systems, exploring perspectives from both users and non-users in home and automotive settings. Semi-structured interviews were conducted with 20 participants with varied experience level with such systems- those who used voice systems in homes, in cars, both contexts and in neither. An inductive thematic analysis revealed key barriers, such as system performance, privacy concerns, passivity, past experiences with similar systems, effort required, user preferences, usage in multi-device settings, and autonomy in using these systems. These findings highlight these key barriers, offering valuable considerations for future design efforts to encourage greater acceptance and effective use of voice systems in both cars and homes.

KEYWORDS

Voice systems, Barriers, Acceptance

Introduction

Voice systems have advanced remarkably in recent years, evolving from basic command-driven models with restricted vocabulary (Heisterkamp, 2001) to sophisticated systems capable of engaging in natural language (McLean & Osei-Frimpong, 2019). Now, with the rise of generative artificial intelligence, these systems can engage in more dynamic and intelligent conversations, offering more immersive and intuitive user experience (Huang et al., 2024; Stappen et al., 2023). In the context of driving, voice systems provide unparalleled convenience by enabling interaction without physical devices, reducing visual-manual distractions (Lee & Jeon, 2022; McLean & Osei-Frimpong, 2019). However, current research on voice systems is primarily focused on home-based technologies, with limited exploration in the driving context (Dutsinma et al., 2022). Moreover, the applications of voice systems in vehicles predominantly revolve around simple tasks such as music control or adjusting functions like temperature and their utilization has not yet garnered widespread acceptance within the driving community (Lee & Jeon, 2022; Luger & Sellen, 2016). In this regard, it is important to note that acceptance of technology hinges largely on subjective opinions of users and cannot be guaranteed just by a well-designed product (Large et al., 2019). Hence, it is essential to explore the factors influencing the use of voice systems in driving contexts from a user-centred perspective, by understanding the views of individuals with diverse mindsets, learning from the experiences of both users and non-users in home and driving settings, and gaining insight into their reasoning behind these preferences.

Method

Semi-structured interviews were conducted with 20 participants (mean age: 30.6) recruited using a purposive sampling technique. The sample size was determined by the exploratory nature of the study. Participants came from four different user backgrounds*: non-users of voice systems, users only at home, users only in cars, and users in both settings*, as shown in Table 1. However, this was not a controlled variable; the aim was to gather perspectives from individuals with diverse usage patterns, and not to conduct statistical comparisons. As anticipated, the group of participants who used voice systems only in cars was smaller, underscoring the focus of this study on the underutilization of voice systems in cars. The non-user group provided insights into why they do not use them or why some discontinued their use (however, these are not specifically differentiated in the results), while users offered perspectives based on their experiences. This approach was informed by firstly existing research highlighting a predominant focus on home-based voice systems, with limited attention to vehicle-specific voice systems (Dutsinma et al., 2022); and secondly, the disparity in usability and lower satisfaction levels with in-vehicle voice systems (Schmidt et al., 2018), thereby prompting an exploration between home-based voice systems and car-based voice systems.

Table 1: Distribution of participants across user groups

Home		Home and Car	Cars	Do not use currently	
Home	Stopped using in cars, using at homes			Never used	Stopped using at homes and cars
P4	P3	P1	P18	P2	P7
P6		P5		P12	P19
P9		P8		P13	
P10				P17	
P11					
P14					
P15					
P16					
P20					
∑ 9	1	3	1	4	2
10		3	1		6

The interviews were designed to explore participants' overall attitudes, opinions, and experiences regarding current voice systems, both generally and specifically in the context of driving, thus, aiming to capture a holistic understanding of the preferences and barriers to use of voice systems among individuals from the different user groups. Interviews were conducted via Microsoft Teams, with each session lasting approximately 30 minutes. The interview recordings were transcribed, and a thematic analysis was conducted using an inductive approach, allowing themes to emerge naturally from the data. Transcripts were reviewed multiple times to ensure familiarity with the content, and relevant segments highlighting barriers (what inhibits their usage?) and motivations (Why are they used?) were noted. These segments were manually coded and collated into overarching themes.

Results

The following 8 themes were identified, reflecting both home and car applications (unless specified in the text), as shown in Figure 1.

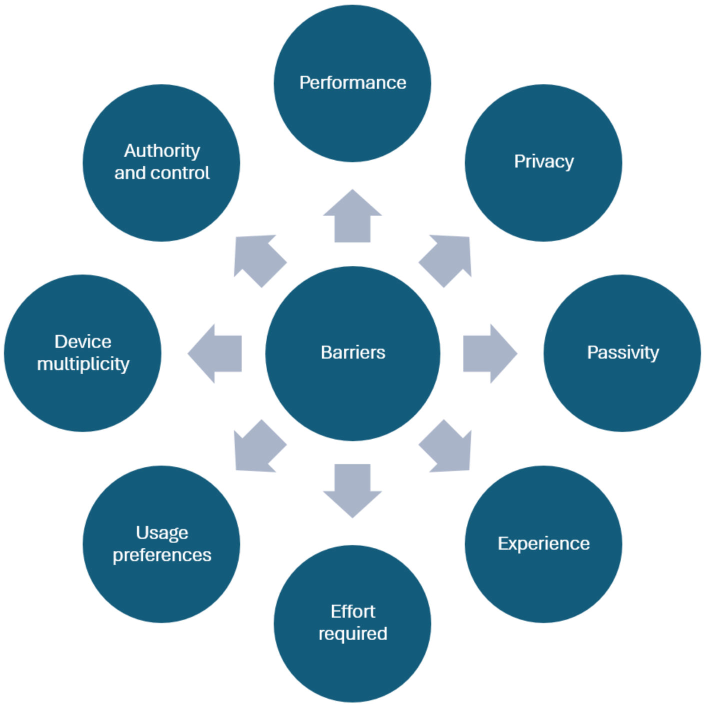


Figure 1: Thematic analysis results

Device performance

Device performance was described as a multifaceted concept, encompassing various elements that together shape the overall perception of the system. Key aspects included response accuracy, recognition accuracy, and wake-word accuracy. **Response accuracy** refers to the correctness of the information provided, or the action performed by the voice system *("I would use it, if I was confident that it would be accurate."* - P18). **Recognition accuracy** pertained to the voice system's proficiency in accurately recognizing and understanding the user input (*"they have issues with them mishearing or umm picking up the wrong thing"* -P8). This subtheme included, but is not limited to, concerns related to diverse accents (*"It's hard for them to understand my accent, so that's frustrating, and it limits my usage"* - P3). In the context of recognition accuracy, concerns were also shared about the ability to capture the command correctly amidst ambient noise in a car such as road noise, music, conversations, and environmental influences (*"depending on the background noise … would be my big worry"* -P10). **Wake word responsiveness** refers to the device's responsiveness to the wake word. It includes situations where the device either gets triggered unintentionally or instances where they fail to activate on the utterance of the wake word. Wake word issues further extended to users preferring a button to trigger the device in cars ("*I'll say the wake word, it won't pick up at all, it won't even respond… I can use wake word in the car, but I don't trust it … when I press the button, I know that it will start listening to me.* -P8). Performance concerns were especially pronounced in the driving context, where a poorly performing system could significantly frustrate drivers, detracting from their focus and potentially compromising overall safety.

Privacy concerns

Alongside performance, privacy concerns were frequently mentioned. Participants' perspectives on privacy could be broadly classified into two groups with differing viewpoints. The first group expressed strong reservations about such systems constantly listening to conversations, fearing they might record private discussions. They believed these systems fostered a sense of surveillance heightening their privacy concerns. Issues were also raised about data being recorded during interactions with young children, unaware of privacy implications, and about targeted advertisements, which led users to believe the systems were monitoring conversations ("*I'm concerned that if the microphones are always on, then it's also recording everything it hears, which means it can tailor adverts*" -P4). On the other hand, the second group believed that the allure of convenience often outweighed their privacy concerns. This group viewed voice systems similarly to mobile phones regarding privacy issues, arguing that they continue to use mobile phones despite knowing that personal data could be shared, ("*I'm okay with the data they collect, I don't think it's different from using the phone itself.*" -P5).

Passivity

Participants indicated that one potential reason for the underutilization of voice systems was the system's 'passivity'. Passivity in this context, refers to the system's state being silent or inactive until triggered by the user. Participants noted instances where they overlooked voice systems, and used mobile phones, as the thought of using the voice system didn't occur to them. This was attributed to the system's dormant nature and also their limited awareness of its capabilities, which likely contributed to reduced usage ("*if I knew their features, it might use them*" -P12), ("*I just do it myself because I forget what it can do*" -P17).

Experiences

In discussing their usage patterns, participants drew upon their experiences with various voice-based devices or applications utilised in the past, such as Amazon Alexa or Apple Siri. Notably, the influence of chatbots and Interactive Voice Response (IVR) systems, particularly those integrated into helpline services, significantly shapes users' perceptions of voice systems and poses a barrier to their widespread adoption and acceptance ("*When I call our service provider, (indicating IVR system) and when I say something it very often gets it wrong*" -P13). In the driving context, participants' experiences were shaped by highly restrictive command-based systems, leading to a diminished user experience and creating a negative impression, consequently undermining their trust in the system. Participants exhibited little interest in using it again considering the consequences to safety (*"... distraction element of it taking away from the safety of driving is probably something I've been frustrated with, like shouting at the car when it wasn't working out what I was saying"* -P19).

Effort

Effort was evaluated based on both the initial setup, the resolution of issues encountered during use, as well as use in general. Participants found the initial setup process complex or adding a new 'feature' or additional functionality (also known as 'skills' on Amazon Alexa) cumbersome, with unnecessary steps often leading to abandonment or disinterest *("... the amount of steps it took to get to it (meaning a game) ... it seemed very long winded. ... by the time it had gone through it, the kids were bored, ... confirmation and validation steps in voice are just prohibitively complicated and long and I've found myself just switching off in the end"* - P20). Additionally, the way the system communicated issues or provided feedback was not perceived as intuitive. Furthermore, the lack of specific error messages made it difficult to diagnose problems, leaving users with the impression that the system had failed without a clear cause, ("*the interface lights up to signify different thing, but apart from acknowledging* *that it's listening. ... I cannot be bothered to remember what light (indicating colour of the light) means what… instead of just using light, perhaps also using some kind of ... not communicating problems faced, so suddenly we just feel that it stopped working and that causes frustration"* - P11).

Usage preference

Participants also expressed a preference for performing tasks manually (for example, changing the heating temperature), using buttons or dials, rather than using a voice system, especially in the context of driving (*"I quite like having tangible buttons in my car, not a huge fan of voice"* -P10); (*"In my car, I've got physical buttons for things like heating ... I can just turn a little wheel. I don't ... think these they need to be replaced."*-P17). They also emphasized that voice systems restrict their ability to use preferred applications, especially for tasks like music playback, where they feel confined to supported apps. Additionally, participants indicated that voice systems require real-time verbal communication, limiting their ability to pause and reflect. In contrast, typing on a mobile phone allows them to compose queries at their own pace, which is why they prefer not to rely on voice systems all the time, (*"If I type something, I know what I'm doing whereas when I talk, it's a bit more woolly and it could be misunderstood"* -P13).

Device multiplicity

This study highlighted challenges associated with the multiple voice system in the users' life. For instance, within a household, the presence of distinct voice interfaces in different rooms led to the incorrect device responded to commands (*"it will just start playing like on the other device and, I don't want it to be on there, someone else has asked for it"* -P20). Concerns were raised regarding the compatibility of voice systems across various settings, including homes, cars, and mobile devices, due to differences in operating systems. Participants expressed a need for enhanced communication capabilities between devices, emphasising the importance of seamless interoperability ("W*hen I'm driving, I struggle to switch between Alexa and Siri, I always end up calling the one in the car Alexa, and then obviously it doesn't do anything because it's Siri"* -P1).

Authority and control

This research emphasized users' desire for authority and control while interacting with the system, as commands can be issued by anyone in the vicinity. Concerns were voiced about children circumventing parental controls through voice commands, with apprehension regarding access to age-inappropriate or explicit content ("*My kids have used it to get around the YouTube blockage*" -P20). Concerns were raised about the safety implications of the possibility of children or passengers issuing voice commands while in transit, as such actions could not only distract but also compromise the overall safety of the journey, (*"If the kids could go, can you reroute us to a toy shop? That would kind of stresses me"* -P17).

Safety concerns

In the driving context, most of above-mentioned issues were directly tied to distraction and then consequences to safety, which was considered as a major barrier while driving ("... *the distraction element of it taking away from the safety of driving is something I've been frustrated with when it didn't work out"* -P19).

Summary and planned further work

The findings of this study highlight the crucial role of device performance in the adoption and usage of voice systems, particularly in driving contexts where performance failures can compromise safety. Key aspects such as response accuracy, voice recognition, and wake word responsiveness must be optimized. While technological advancements enhance performance, voice systems should support users in forming accurate mental models for effective interaction. A well-formed mental model improves user understanding of system capabilities and limitations, enabling seamless human-computer interaction (Razin et al., 2021). Privacy, security, and ethical concerns also emerge as significant barriers to acceptance, necessitating policymakers, regulators, and industry stakeholders to ensure transparency regarding data collection, storage, and access permissions as suggested by (Gumusel et al., 2024; Lau et al., 2018). Additionally, users often lack awareness of system features which emphasizing the need to showcase capabilities without becoming intrusive, by implementing relevant marketing strategies and through proactive system design (Luger & Sellen, 2016). Efforts should be made to reduce setup and troubleshooting issues, as minimal feedback can lead users to perceive system failure. This research suggests that voice systems should be designed to allow easy integration of preferred applications. Moreover, this research also highlights new system behaviours should be introduced as optional features that users can disable if needed. This aligns with users' preference for having control and authority over the device and their need for it to be flexible enough to suit their needs. The study also highlights confusion arising from interactions with multiple voice systems. Lastly, negative experiences with older command-based systems persist, indicating a need for strategies to reframe user perceptions and encourage the use with advanced voice systems through further research.

These insights provide an initial understanding of the barriers that exist in the acceptance of voice systems. These findings will be used to develop a novel voice system acceptance questionnaire, designed to capture large-sample data. This will be used to compare perceptions of voice systems in the home versus in the car and analyse these differences to generate design guidance to remove/reduce barriers to in-vehicle acceptance.

References

Dutsinma, F. L. I., Pal, D., Funilkul, S., & Chan, J. H. (2022). A Systematic Review of Voice Assistant Usability: An ISO 9241–11 Approach. *SN Computer Science*, *3*(4). https://doi.org/10.1007/s42979-022-01172-3

Gumusel, E., Zhou, K. Z., & Sanfilippo, M. R. (2024). *User Privacy Harms and Risks in Conversational AI: A Proposed Framework*. http://arxiv.org/abs/2402.09716

Heisterkamp, P. (2001). *Linguatronic product-level speech system for Mercedes-Benz cars*. https://doi.org/10.3115/1072133.1072199

Huang, S., Zhao, X., Wei, D., Song, X., & Sun, Y. (2024, May 11). Chatbot and Fatigued Driver: Exploring the Use of LLM-Based Voice Assistants for Driving Fatigue. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3613905.3651031

Large, D. R., Harrington, K., Burnett, G., Luton, J., Thomas, P., & Bennett, P. (2019). To please in a pod: Employing an anthropomorphic agent-interlocutor to enhance trust and user experience in an autonomous, self-driving vehicle. *Proceedings - 11th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2019*, 49–59. https://doi.org/10.1145/3342197.3344545

Lau, J., Zimmerman, B., & Schaub, F. (2018). Alexa, are you listening? Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW). https://doi.org/10.1145/3274371

Lee, S. C., & Jeon, M. (2022). A systematic review of functions and design features of in-vehicle agents. *International Journal of Human-Computer Studies*, *165*, 102864. https://doi.org/10.1016/J.IJHCS.2022.102864

Luger, E., & Sellen, A. (2016). “Like having a really bad pa”: The gulf between user expectation and experience of conversational agents. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/2858036.2858288

McLean, G., & Osei-Frimpong, K. (2019). Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior*, *99*. https://doi.org/10.1016/j.chb.2019.05.009

Razin, Y. S., Gale, J., Fan, J., Smith, J., & Feigh, K. M. (2021). WATCH FOR FAILING OBJECTS: WHAT INAPPROPRIATE COMPLIANCE REVEALS ABOUT SHARED MENTAL MODELS IN AUTONOMOUS CARS. *Proceedings of the Human Factors and Ergonomics Society*, *65*(1). https://doi.org/10.1177/1071181321651081

Schmidt, M., Braunger, P., Sartor, J. B., Hondt, T. D. ', & De, W. (2018). *A Survey on Different Means of Personalized Dialog Output for an Adaptive Personal Assistant*. https://doi.org/10.1145/3213586.3226198

Stappen, L., Dillmann, J., Striegel, S., Vögel, H. J., Flores-Herr, N., & Schuller, B. W. (2023). Integrating Generative Artificial Intelligence in Intelligent Vehicle Systems. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*. https://doi.org/10.1109/ITSC57777.2023.10422003

Self-Assessing Visual Attention While Driving: Implications for Driver Monitoring Systems

Max Harding[1], Catherine Harvey[1], David R. Large[1], Gary Burnett[2], Chrisminder Hare[3] & Karl Proctor[3]

[1]University of Nottingham, [2]Loughborough University, [3]Jaguar Land Rover

SUMMARY

Driver Monitoring Systems (DMS) are now a requirement in all new vehicles. DMS aim to reduce crashes and improve driver attention by providing warnings or interventions to the driver. However, driver acceptance is crucial to ensuring their effectiveness. Various methods exist to detect inattention, but if these do not align with a driver's mental model, acceptance issues may arise regarding the warnings provided. The study outlined in this paper examines how drivers assess their own visual attention by comparing self-reported ratings to a visual attention algorithm. Using an on-road experiment with six participants, initial results highlight the importance of contextual information for accurately assessing visual attention and providing effective warning strategies.

KEYWORDS

Driver Monitoring, Driver Attention, Visual Inattention, Driver Acceptance

Introduction

Systems monitoring driver attention are a requirement for all new vehicles via the General Safety Regulations (GSR) of the European Union (European Commission, 2023). Similarly, Euro NCAP details Driver Monitoring Systems (DMS) requirements and guidelines for using DMS with driving assistance features (Euro NCAP, 2023, 2024). The key aim of DMS is to reduce the number of crashes related to visual inattention by providing warnings to redirect a driver's attention back to the driving task, or through other interventions, such as breaking or disabling driver assistance features. It is a GSR requirement that the driver can deactivate these systems. Driver acceptance could be affected if the system is intrusive or prone to false alarms, so it should only warn drivers when genuine inattention is detected (Smyth *et al.*, 2021). There are various methods for estimating driver attention with strong anchors in driver attention theory (Ahlström, Georgoulas and Kircher, 2022). However, attempting to define a ground truth to compare visual attention algorithms against is a difficult task. Nevertheless, the correspondence between DMS outputs and a driver's perception of their attention will likely impact acceptance.

The study looks at a novel on-road method comparing how drivers self-assess their visual attention against the output of the visual attention algorithm AttenD (Kircher and Ahlström, 2013). The output of the AttenD algorithm is a buffer value between 0 and 2, which changes over time depending on where the driver is looking. The buffer increases in value with glances to the road and decreases with glances away from the road or when the driver is visually distracted. Drivers are given additional time when performing driving related tasks such as checking mirrors or speed, before the buffer begins to decrease. When the value of the AttenD buffer drops to 0, the driver is considered visually inattentive. During the study, we examined cases in on-road driving where ratings of inattention by the AttenD algorithm and the driver differed. These discrepancies helped explore factors that influence a driver's assessment of their attention and examine scenarios where DMS acceptance may be impacted.

Methods

Six participants (3 male, 3 female) were recruited for an on-road study comprising a 20-minute drive. To encourage naturalistic driving behaviour, participants could follow a familiar route or use a satnav for navigation if unfamiliar with the local area. Participants wore SMI eye-tracking glasses and drove in the research institute's Nissan Leaf Test Vehicle. Before the drive, participants were briefed on the definition of visual attention and were given examples of attentive and inattentive visual behaviour while driving. Participants were instructed to drive as they normally would. At 30-second intervals, a short audible tone prompted participants to rate their visual attention in the period since the last prompt. Participants were instructed to verbally respond with "yes" if they felt they had been visually attentive since the last prompt, or "no" if they felt they had been visually inattentive, followed by a short reason if it was safe. Other rating scales, such as Likert scales, were piloted. However, the complexity of the scale made it difficult for drivers to distinguish between visual and cognitive attention when giving ratings.

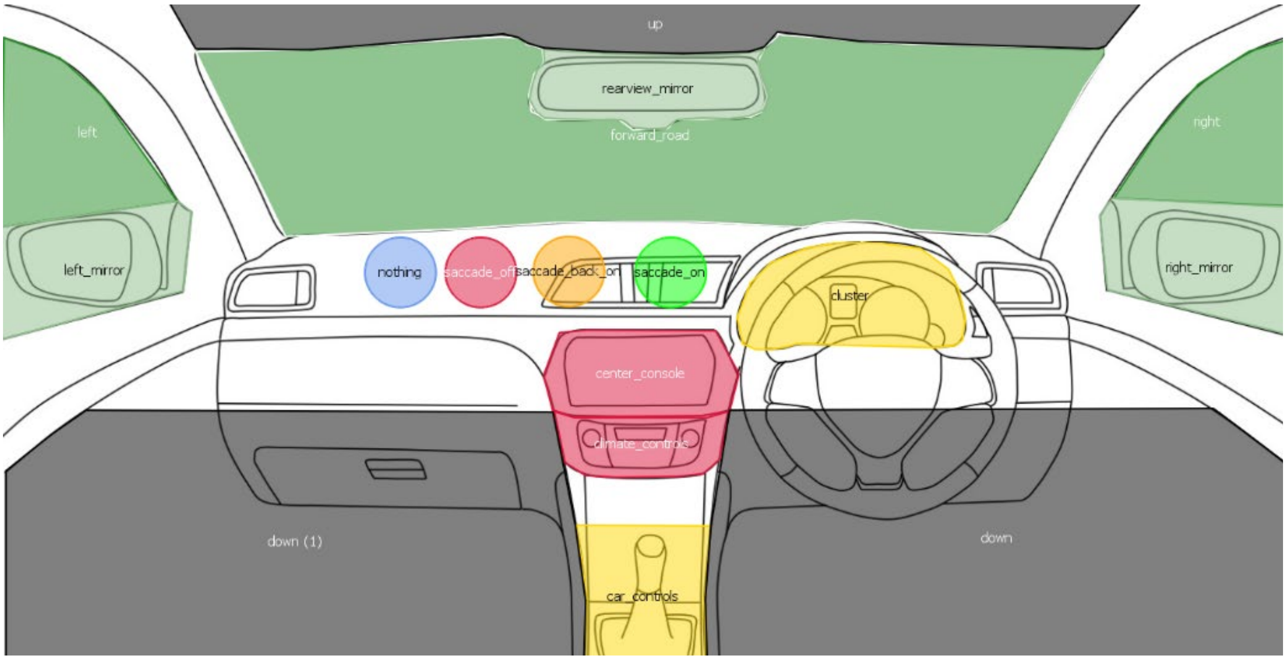


Figure 1: Areas of Interest for semantic gaze mapping

Following each session, the eye tracking data was analysed using semantic gaze mapping to map fixations and saccades to areas of interest defined by AttenD as shown in Figure 1. The AttenD algorithm was retrospectively applied to the gaze data to generate an attention profile across each drive. To compare the participant's self-assessments to the AttenD algorithm, the sections across which drivers were prompted to rate their visual attention, were overlaid on top of the output of the AttenD algorithm. If the AttenD buffer reached zero in any of these sections, then the AttenD rating for that section was marked as inattentive. This simplified output of the AttenD buffer can be easily compared to the driver's binary rating scale ("yes/no").

After the analysis of the gaze data, and processing of the data using AttenD, participants were invited back to review clips of their eye tracking from the drive. An example frame from a clip is shown in Figure 2. Participants were presented with clips where there was either a disagreement between their own rating and the output of the AttenD algorithm, or where both they and the AttenD algorithm had given a rating of inattentive. Participants did not have knowledge of their initial ratings or the output of the AttenD algorithm while watching the clips. Participants were asked to reassess their visual attention based on the clips to explore whether their post-drive responses aligned with their responses during the drive. The time frame between the initial drive and the participant's interviews was kept to a minimum (typically a few days) so that the drive was still familiar to the participant. After the participants reassessed each clip, they were informed of their initial rating and the AttenD algorithm's rating. Participants were then asked whether they would have been receptive to an 'intervention' or prompt at that point in time, alerting them that their visual attention was low in that scenario. Participants were also asked what type of warning they would prefer/expect in that scenario.

Figure 2: Frame from the recording of a participant's eye tracking.

Results and Discussion

Comparison of Participant Ratings

In total, there were 256 ratings of attention across the 6 participants. The results of the study are summarised in Table 1. For the purpose of this discussion, AttenD is taken as the ground truth (i.e. AttenD is a true measure of the participant's visual attention. The data in Table 1 are the signal detection theory comparisons of 'True Positive' or 'Hit', 'True Negative' or 'Correct Rejection', 'False Negative' or 'Miss' and 'False Positive' or 'False Alarm' (Stanton *et al.*, 2009). Table 1 shows that some participant's assessments changed between the drive and post-drive interview. Out of the ratings given during the drive 79% matched the algorithm, this increased to 92% post-drive. Two of the six participants did not change their responses between the drive and the interview. Changes in false positive outcomes were due to participants amending their initial assessment of inattentive to attentive upon post-drive review (note that 70% of these changes are attributed to one participant, who stated their ratings were overly cautious during the drive). For the majority of changes to False Positive ratings, participants could not see a reason for their original rating. For two of the False Positive ratings remaining post-drive, the participant stated they had been staring ahead and daydreaming. As they were looking ahead, the algorithm would not have been able to pick up on this. The other two remaining false positives were when the algorithm got very low but not quite to zero and a case where a driver felt they were still visually inattentive.

Table 1 - Participant Assessment vs AttenD Ratings

Outcomes	Drive	Post-Drive	Description
True Positive	17	24	Both AttenD and Participants Rated Inattentive
True Negative	181	207	Both AttenD and Participants Rated Attentive
False Positive	30	4	AttenD Rated Attentive, Participant Rated Inattentive
False Negative	28	21	AttenD Rated Inattentive, Participant Rated Attentive

Similarly, the post-drive reduction in False Negative outcomes is due to participants changing their attentive ratings to inattentive ratings on review. Of the remaining False Negative ratings after the post-drive interview, 66% of these occurred while participants were stopped at or approaching traffic lights. This suggests that there is a change in what the participants perceived should be classed as inattentive behaviour depending on the scenario. The remaining 34% of disagreements were due to drivers simply disagreeing that they were inattentive while driving. The discrepancy with AttenD and changes in their own assessment upon review highlights drivers' difficulty in assessing their visual attention in the moment, or differences in understanding of what visual attention is. There were also 5 cases due to known issues with the AttenD algorithm, where a false rating of inattentive was given. The false rating occurred when drivers looked away from the forward roadway for extended periods for genuine reasons, such as checking right to see if a roundabout or junction was clear. Without contextual information about where the driver's glances should be prioritised, the algorithm cannot accurately assess the driver's visual attention in all scenarios.

Participants Responses to Warnings

The study also looked at participants' opinions on warnings in response to a DMS detecting visual inattention. The False Negatives due to known issues with AttenD, were excluded from the analysis. Naturally, participants said they would not be open to warnings for the False Negative outcomes where they had rated themselves as visually attentive. Within the True Positive outcomes after the interview, participants said they would not be receptive to warnings in 30% of cases. The times when participants said they would not be receptive to interventions within the True Positive Outcomes, were similar to the cases they had disagreed with AttenD in that they mainly occurred when the participants were stopped or stopping.

Overall, participants said they would not be receptive to warnings in 57.5% of the cases where AttenD rated inattentive (irrespective of how the participant rated). These responses demonstrate the importance of the output of a driver attention algorithm aligning with the drivers' mental models for acceptance of DMS. Looking closely at the clips where participants were not receptive to warnings, 57% occurred while driving, and 43% occurred when participants were stopped or approaching traffic lights. The clips where drivers stopped at traffic lights point to a difference in what the participant and AttenD consider inattention. These cases showcase that contextual information is important for accurately detecting inattention (Kircher and Ahlström, 2024) and acceptance of warnings. Participants also said they would be more accepting of specific warnings in different scenarios and would want the modality or severity of the warning to change depending on different factors. The results highlight the need for contextual information to be accounted for future DMSs in determining driver attention and the modality that warnings about their attention are given.

Conclusion

The study explored the differences between driver self-assessments of visual attention and the AttenD algorithm's outputs. The findings of the study highlight the importance of contextual information for accurately assessing visual attention and raise questions about how visual attention should be assessed in different scenarios. Participants often revised their self-assessments on review post-drive, revealing difficulties in accurately assessing their visual attention in real time, which could further contribute to acceptance issues. The study also highlighted that participants said they would be less receptive to warnings in scenarios where they did not perceive themselves as being visually inattentive. Even when participants agreed they were visually inattentive, they would still not have been receptive to some warnings in specific scenarios, such as being stopped at traffic lights. The discrepancies between AttenD outputs and driver self-assessments suggest that DMS should integrate contextual awareness to better align with the driver's mental models. Further research will explore the impact on driver acceptance of different modalities and forms of warnings from DMS in different scenarios.

References

Ahlström, C., Georgoulas, G. and Kircher, K. (2022) 'Towards a Context-Dependent Multi-Buffer Driver Distraction Detection Algorithm', *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, 25(5).

Euro NCAP (2023) 'Assessment Protocol – Safety Assist Safe Driving'.

Euro NCAP (2024) 'Assisted Driving - Test and Assessment Protocol'.

European Commission (2023) 'Commission Delegated Regulation (EU) 2023/2590 of 13 July 2023', *Official Journal of the European Union*, (L 312), pp. 1–15.

Kircher, K. and Ahlström, C. (2013) 'Chapter 19 The Driver Distraction Detection Algorithm AttenD', in *Driver Distraction and Inattention: Advances in Research and Countermeasures, Volume 1*. SAE, pp. 327–348.

Kircher, K. and Ahlström, C. (2024) 'A comparison of glance coding approaches for driver attention assessment', *Transportation Research Part F: Traffic Psychology and Behaviour*, 100, pp. 243–253. Available at: https://doi.org/10.1016/j.trf.2023.12.003.

Smyth, J. *et al.* (2021) 'Public acceptance of driver state monitoring for automated vehicles: Applying the UTAUT framework', *Transportation Research Part F: Traffic Psychology and Behaviour*, 83, pp. 179–191. Available at: https://doi.org/10.1016/j.trf.2021.10.003.

Stanton, N.A. *et al.* (2009) 'Predicting pilot error: Testing a new methodology and a multi-methods and analysts approach', *Applied Ergonomics*, 40(3), pp. 464–471. Available at: https://doi.org/10.1016/j.apergo.2008.10.005.

An electric aircraft accident case study – human factors analysis and lesson learned

Michael A. Bromfield[1,2]

[1]University of Birmingham,[2]Myriad Business Ltd

SUMMARY

An electrically powered aircraft was conducting flight tests, under experimental flight rules when power to the electrical motors was lost. The two-person crew took 9 seconds to identify the failed system and attempt recovery actions which were unsuccessful and the pilot landed in a field adjacent to the airport resulting in severe damage to the aircraft. HFACs was used to identify causal and contributory factors and highlight key learning points. The results informed changes to UK CAA experimental flight rules related to flight test organisation, use of system controls, positioning of displays and the availability of alerts and warnings.

KEYWORDS

Accident investigation, electric aircraft, human factors, regulation, flight test

Introduction – the accident

The electrically powered aircraft conducting flight tests under E Conditions (CAA, 2019) experienced a loss of power during an interruption of the power supply (AAIB, 2022). During this interruption the windmilling propeller generated a voltage high sufficient to lock out the power to the motors. The pilot and observer were unable to reset the system and restore electrical power and the pilot made a forced landing. The aim of this research is to identify possible/probable causal and contributory factors and highlight key learning points to inform future design, development and flight testing of electric or hybrid-electric aircraft.

Method

An in-depth review of the aircraft accident report (AAIB, 2022) was conducted using the Human Factors Analysis and Classification System (Wiegman & Shappell, 2002). HFACS is commonly used in aircraft accident investigation and the author has extended it to include social, economic or political factors (latent outside influences) and aircraft handling qualities and certification (active or latent preconditions for unsafe acts). Guidelines for the design and evaluation of flight deck displays and controls (FAA, 2016) and cockpit assessments (Qinetiq, 2015) were used to assess the cockpit environment. Regulation for the operation of aircraft under experimental conditions at the time of the accident was reviewed (CAA, 2019).

Results

An independent review of the accident report (AAIB, 2019) and the application of HFACS by the author yielded the following contributory factors within the extended HFACS framework.

Outside influences (latent)

The pressure to be 'first to market' and demonstrate flight with a zero-emission aircraft were likely to be present (social). The company received external grant funding and were committed to deliver on-time within the allocated budget. The flight test programme was delayed and this may have led to additional internal (unfunded) costs (economic/commercial pressures).

Organisational influences (latent)

Most specialist staff were from outside of the aviation industry. The competent person responsible for flight testing, lacked formal flight test education, training and experience. Flight test observer and flight test director lacked flight test education, training and experience (resource management). Staff were passionate and highly motivated working in a fast-paced and goal-oriented problem solving environment. Safety was not talked about as a key priority, essential in a flight testing environment (culture). There were few prescribed procedures and staff were given autonomy to develop practical workarounds or adaptions to solve problems and overcome constraints. There was a lack of independent risk assessment of the flight test programme and lack of post-flight debrief including 'lesson learned' (organisational process).

Unsafe supervision (latent)

The competent person responsible for flight testing was not actively engaged in flight test activities. The competent person did not enlist the help of other individuals with the appropriate knowledge, skills and experience for a project of this complexity (inadequate supervision). There was inappropriate test planning and organisation and lack of guidance on functional links between individuals in charge of flight test activity or how coordination was to be achieved. 'Plan to fly' and 'fly to plan' flight test principles were not followed (planned inappropriate operations). The propulsion system had failed on two previous flights however the risk assessment was not updated (failed to correct problem).

Preconditions for Unsafe Acts (active or latent)

Operational constraints at the airfield provided limited options for a forced landing. Poor design and positioning of the propulsion system management displays resulted in delays to identify/confirm situation. Blanking out of selected legacy instruments required adaption of pilot instrument scanning. A step to retard power lever in Aircraft Flight Manual (AFM) emergency procedures was missing. Propeller speed was controlled by rotary control fitted to side of power level quadrant (unconventional) (environmental factors/technological environment). There was a lack of Crew Resource Management/Threat & Error Management training in a flight test environment. The multi-crew flight test environment was treated as a single pilot operation with an observer (personnel factors/Crew Resource Management). The landing performance of the aircraft was not considered by the pilot. The aircraft was loaded above maximum take-off weight for the basic (unmodified) aircraft and 60 lb below the E-conditions limit. This meant that the aircraft was 'heavy' to fly resulting in degraded climb performance (condition of aircraft/handling qualities).

Unsafe Acts/Errors (active)

The pilot did not call for the emergency checklist. Multiple attempts were made to restart resulting in degraded flight path management. Flaps and landing gear lowered late (errors/decision errors). The aircraft deviated from heading on the downwind leg, deviated from height during base and finals then misjudged the forced landing (errors/skill-based). The lack of situation awareness during the propulsion system troubleshooting resulted in an overshoot of the runway (errors/perceptual).

Summary

‘First to market’ and commercial pressures led to workarounds and adaptions. Attention to safety and risk assessment were compromised due to lack of formal flight test education. The competent person was not an integral part of the flight test team (CRM). The design of the human-machine interface was sub-optimal for a high-risk test environment, resulting in reduced situation awareness for both pilot and observer. Procedural errors and delays led to deviation from the intended circuit pattern and approach profile.

Conclusions

This case study has highlighted the importance of an integrated approach to the application of human factors principles to design and development in a high risk, flight test environment where crew training and a systems thinking approach are essential for safe outcomes. Appropriate knowledgeable, skilled and experienced resources are essential to support technically challenging programmes. Regulatory oversight needs to be measured to keep pace with technological advancement but not stifle innovation. As a result of this analysis and recommendations made in the AAIB report (AAIB, 2022), changes to the E Conditions regulation were made with respect to the design of cockpit displays, controls, alert and warning systems (CAA, 2024).

References

AAIB, (2022). G-HYZA, Piper PA-46-350P Accident Report AAIB-27260, Crown Copyright, 2022. https://www.gov.uk/aaib-reports/aaib-investigation-to-piper-pa-46-350p-modified-g-hyza

CAA, (2019). Operation of experimental aircraft under E Conditions, CAP 1220, Edition 2, November 2019.

CAA, (2024). Operation of experimental aircraft under E Conditions, CAP 1220, Edition 3, November 2024. https://www.caa.co.uk/our-work/publications/documents/content/cap1220/

FAA (2016), Human Factors Considerations in the Design and Evaluation of Flight Deck Displays and Control, version 2, DOT/FAA/TC-16/56, December 2016. pp183~230 https://rosap.ntl.bts.gov/view/dot/12411

Qinetiq, (2015). Cockpit Assessment, Section 1.12, ETPS Flight Test Manual, Version 1.0, Qinetiq Ltd, 2015.

Wiegman, D.A, & Shappell, S.A, (2000). Human Factors Analysis & Classification System. DOT/FAA/AM-00/7 "The Human Factors Analysis and Classification System - HFACS"

Anatomy of a Deep Dive into Airport Taxiway Incidents

Kirwan, B[1], Newman, D[2], Elliott, R[3], Biliri, E[4] and Bettignies-Thiebaux, B[1]

[1]EUROCONTROL, [2]National Air Traffic Services, [3]London Luton Airport, [4] Suite 5

SUMMARY

A Deep Dive exercise was carried out at London Luton Airport (LTN), exploring taxiway incidents with four airlines, Air Traffic Control (ATC), the airport authority London Luton Airport (LLA), a Human Factors expert and a data scientist. The exercise focused on developing a better understanding of Taxiway Errors at LTN and identifying potential mitigations. Eleven key factors underpinning taxiway errors, and nine mitigation pathways were identified, several of which are now being implemented. This paper illustrates the Deep Dive process, and the resources required to make it a success, and recommends it as an agile yet straightforward technique for the Human Factors practitioner.

KEYWORDS

Airport Safety, Deep Dive, Incident Analysis

Background

London Luton Airport (LTN) is the UK's fifth busiest airport. Although LTN has an excellent safety record, taxiway errors continue to occur, albeit on an infrequent basis as they do at many airports. Figure 1 shows the single runway airport and taxiway layout at LTN. Taxiway errors can occur for example when the flight crew of an aircraft become confused about where they are on the taxiway system with respect to other aircraft, or where they need to go. This can result in potential encounters with other aircraft. Whilst such encounters are detectable, and often result simply in delays (e.g. as one aircraft may have to back-track), they can significantly increase workload, and ultimately, they carry safety risk. Analysis of several years' worth of such incidents – which tend to be rare and often seem to have a unique set of causal or contributory factors – did not reveal many specific patterns or candidate countermeasures to reduce the risk of occurrence. Therefore, as part of the Luton Airport (LTN) Safety Stack's ongoing safety programme, a Deep Dive exercise was carried out, lasting a single day, focusing on developing a better understanding of Taxiway errors at LTN, and identifying associated mitigations.

A deep dive (see Kirwan et al, 2021) brings together diverse expert knowledge and attempts to view the problem from multiple angles and aims to go beyond traditional assumptions about what safety barriers – whether physical, procedural or cultural – are working. The following list captures the aims and analytical avenues explored during a deep dive:

- Explore a specific accident or incident trend
- Examine the basis for safety
- Which barriers are working?
- Which barriers are no longer working?
- What are the key Human Factors involved (both positive and negative?)
- Have any external factors changed?

- Have internal factors changed (staffing, training, etc.)?
- Are the procedures still fit for purpose?
- What are the deeper systemic factors?
- Where are the hotspots?
- Where are there best practices?
- What can be shared across airlines & the Stack?

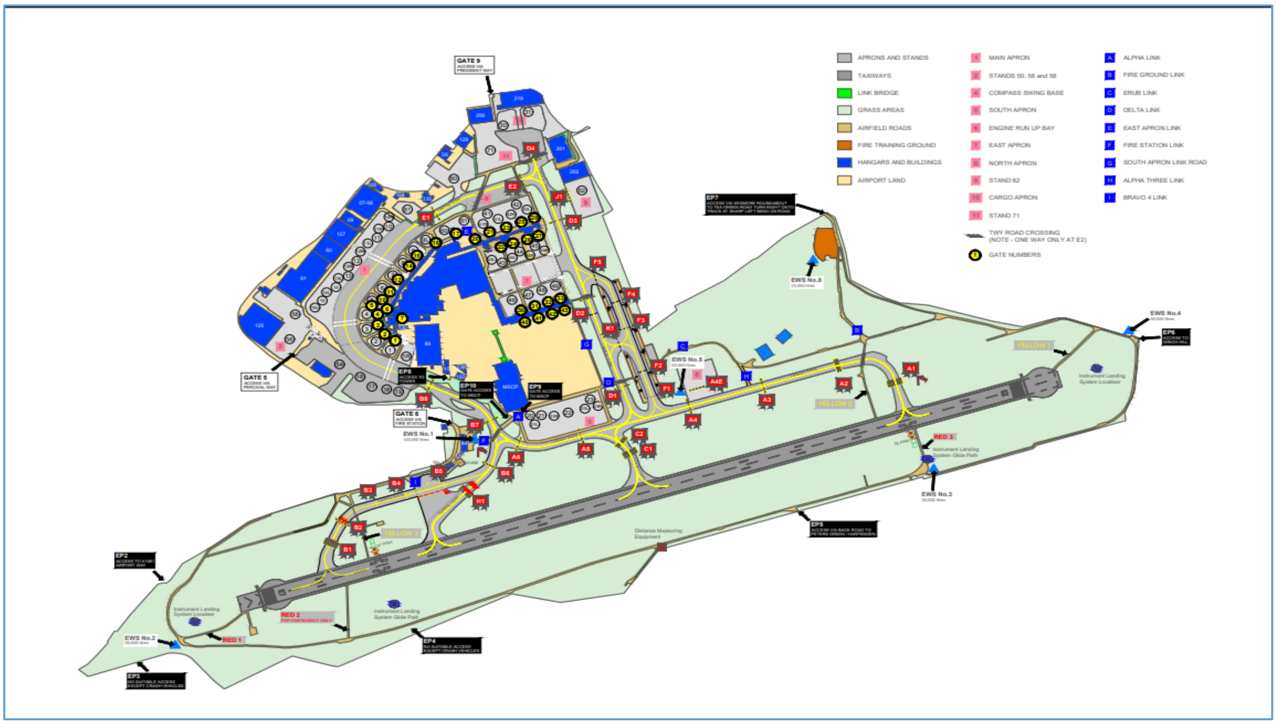


Figure 1: Airport and taxiway layout at London Luton Airport

The Deep Dive was run by EUROCONTROL and LLA with participation from the air traffic service provider and four airlines (including a business jet company), and a data scientist from the HAIKU project[1] involved in the analysis of such incidents. There were three aims of the exercise:

1. Identify the most important factors driving incorrect taxiway errors
2. Identify potential mitigations
3. Consider ways to improve quality and quantity of reporting

Approach & Findings

The in-presence deep dive took place in early October 2024, in National Air Traffic Services (NATS) premises at the airport. Ten incidents (post-COVID, when traffic levels had re-normalised) were selected by the airport Air Traffic Control provider (NATS), to encompass a range of scenarios including inbounds (arrivals) and outbounds (departures), commercial airlines and business jets, different times of day/year, and different causal and contributory factors. The incident information for each event varied, but it helped that both NATS and LLA had reports on the same incident, and some experts present were well aware of the incidents and their background factors.

Participants were experienced (between ten and thirty years), including a control tower supervisor and the chief incident investigator for LTN Tower, an airline Base Captain, a Safety Pilot from another airline, and two experienced captains from a third airline and a business jet operator.

[1] HAIKU is a European Commission funded project: https://haikuproject.eu

Additionally, two London Luton Airport safety officers participated, as well as a Human Factors and Safety expert from EUROCONTROL, a second EUROCONTROL expert familiar with taxiway / runway incursion incidents from the US, and a data scientist who has helped develop the Safety Dashboard for LTN, which includes the past seven years of taxiway incidents.

After a brief introduction to the exercise aims, and a review of the HAIKU Safety Dashboard with respect to Taxiway Errors (see Figure 2), NATS introduced the first incident. From a deep dive perspective this makes sense, as the tower controllers literally have a birds-eye view of such events and are the ones issuing instructions to pilots on where to go and when to stop. Other participants see the event from closer quarters (e.g. pilots), and LLA have a broader systemic viewpoint as they oversee all airport operations.

All participants then freely discussed the event and its potential causes and contributing factors, as well as potential mitigations to prevent recurrence. Human Factors involvement was key here, to help unpack some of the contributory factors (e.g. fatigue, situation awareness, complacency, memory failures and cultural factors, etc.) and clarify their likely independent or compound impacts on human performance. Human Factors can help avoid the ‘Tower of Babel’ effect where people might use different terms that effectively mean the same thing, as well as help develop a list of valid contributory Human Factors that can then be used to consider countermeasures.

For five of the incidents, it was possible to review them on the Safety Dashboard, though in several cases (for outbounds) there was scant information on why the incident had occurred. The discussions continued in series and benefitted from having pilots who regularly flew from LTN, as well as the NATS Chief Investigator for LTN, who could sometimes provide a little extra insight into factors based on communication with the airlines (i.e. more detail than was in the reports). At lunchtime the controllers escorted the group to the Tower, to show them the ‘birds-eye’ view and give them a visual reference to some of the incidents, and the taxiway layout. After lunch the incident review continued until approximately 15:30, the deep dive lasting six hours in total.

Figure 2: Extract from Safety Dashboard at LTN Airport

Results

Table 1 summarises the contributory factors identified during the exercise, in the vernacular or common parlance of the aviation experts present. This rich contextual description was preserved because the results were to be presented to the LTN Safety Stack community including other pilots, who would instantly recognise the issues.

Table 1: Causal / Contributory Factors identified in the 10-incident review

Air Traffic Controllers	Flight Crew (Pilots)
• Can't expect pilots to identify biz [business] jets • Visual perspective very different between TWR [Tower] and cockpit – ATCOs (Air Traffic Controllers) need to understand the pilot's perspective • Conditional clearances may have reduced HPBs [Hold-Point Busts] but may increase incorrect taxiway selection • Situation awareness • Signage (+ poor vis?) • OJTI [On Job Training Instruction] • Position Handover	• Don't assume, ask • Is wireless pilot the one driving? [Bizjet] • Cultural bias – wants to go first; see conditional clearance as flexible? [General Aviation or business jet issue] Cultural differences in risk understanding or risk norms at other airports. • Language (Pilots whose first language is not English) • Distractions (taxiway looks clear, get distracted by other small tasks, forget to turn) • RWY [Runway] change – familiarity can mean you go left instead of right • Unfamiliarity with airfield • Expectation bias – unusual routing (also if 2nd sector coming back to LTN, will expect same routing) • Construction ongoing (can obscure or confuse viewpoint from the cockpit) • Drivers on mobile phones • Chart size [too small] on some bizjets in the cockpit

It should be noted that one of the above was already seen as improved since these incidents had occurred, e.g. improved signage on the Taxiway system. Others were seen as noteworthy but not having a major impact, e.g. OJTI (On Job Training Instruction) and position handover (both in ATC). Others, e.g. over-familiarity and conditional clearances (e.g. move to this point, then when aircraft XX has crossed, proceed), were viewed as 'double-edge swords', e.g., unfamiliarity with the airfield can lead to poor situation awareness, but over-familiarity can lead to expectation bias or mistaking left for right (or vice versa) when the prevailing RWY direction (25) is reversed (07). Table 2 shows the initial countermeasures identified.

Table 2: Potential Remedial Measures Identified

ATCO	Pilot
• Invite pilots to TWR • Use of 'give way' in ATCO instruction to pilot. • If they [the aircraft] don't stop, bring up a Stop Bar • Use 'After' in the instruction to pilots. • Markings on the ground [e.g. at position Foxtrot on the taxiway system] • Give extra context in clearance (e.g. opposing traffic)	• Take ATCOs in jump seat [so they can better understand the pilot's perspective] • Invite ATCOs into bizjet, to see their perspective. • Type taxi-route into 'scratchpad' in the cockpit before brakes off (so that both pilots in the cockpit are 'on same page') • The airlines all use their own investigation taxonomies – perhaps these could be pooled • Cockpit self-briefings on taxiway route before landing

• Give less context if it can be misconstrued • Emphasise if non-usual routing (e.g. turn *LEFT*) to Stand XY • Enhance demarcation markings on 71L/R • Survey on Taxi phraseology	• Threat and Error Management (TEM) briefings; use of sterile cockpit procedures [no extraneous communication in the cockpit] when on approach until engine power down at gate. • Mechanism for saying 'I don't understand' when pilots are unclear about the ATCO's instruction [rather than standard phraseology 'say again', which usually means 'I didn't hear you, please repeat', rather than 'I heard you but I still don't understand']

The group then took part in two voting rounds. The first was to choose the most important contributory / causal factors, and the second was to select the most useful mitigations identified. In each round, each participant chose their top three factors, and then their top three mitigations (the two sets did not have to be related). The top factors and mitigations were then identified simply as those with the highest score, leading to a 'Top 11' factors and 'Top 9' mitigations. These are shown below, in priority order from top to bottom.

1. Don't assume, ask
2. Distractions
3. Expectation
4. Situation awareness
5. ATCOs need to understand pilot's perspective
6. Conditional clearances
7. Cultural bias – 'wants to go first' / differences in risk understanding
8. Language issues (not everyone's first language is English)
9. Construction (e.g. causing obstructed views)
10. Can't expect pilots to identify business jets by their model
11. Drivers using mobile phones

It is notable that many of these relate to situation awareness. They do not happen to everyone all the time, rather to some rarely. But collectively they add up to a consistent string of taxiway errors and a 'stubborn' incident pattern.

The top nine mitigations are listed below, again, in priority order.

1. Mechanism for saying 'I don't understand'
2. ATCO emphasise if non-usual routing (e.g. turn LEFT) to Stand
3. Enhance demarcation markings on 71L/R
4. If the aircraft doesn't stop when instructed to do so, bring up a Stop Bar
5. Markings on the ground (e.g. as has been done for *Foxtrot*)
6. Survey on Taxi Phraseology to see if it can be made clearer
7. Take ATCOs in jump seat
8. Invite pilots to TWR
9. Type taxi-route into scratchpad before brakes-off (both on same page)

Impact of the Deep Dive

A presentation on the deep dive was given to around 25 companies at the quarterly LTN Safety Stack meeting later in October, and various representatives undertook actions to explore and address the implementation of the remedial measures. Progress was reviewed at the subsequent Stack meeting in January 2025, and is summarised in Table 3 below.

Table 3: Progress on Countermeasures

Countermeasure	Progress
Mechanism for saying 'I don't understand'	Certain airlines have briefed their pilots about this (some already do it). One Base Captain commented that if any of their pilots do not understand, they stop the aircraft until the action is clarified and understood.
ATCO emphasise if non-usual routing (e.g. turn LEFT) to Stand	This practice is being trialled in the TWR by some of the ATCOs.
Enhance demarcation markings on 71L/R	Done.
If the aircraft doesn't stop when instructed to do so, bring up a Stop Bar	Stop Bars are not always useful in such situations, as by the time the controller detects the issue the aircraft may have already passed the Stop Bars. This is still under review.
Markings on the ground (e.g. as has been done for *Foxtrot*)	Done.
Survey on Taxi Phraseology to see if it can be made clearer	Ongoing.
Take ATCOs in jump seat	This has begun and will be further taken up.
Invite pilots to TWR	This has been happening with a number of pilots and will continue.
Type taxi-route into scratchpad before brakes-off (both on same page)	This best practice is already in place by one or two airlines and being considered by others (one Base Captain stated that this should be the top recommendation).

Generalisability of the results

Although it is often said in the industry that 'if you've seen one airport, you've seen one airport', due to different and often unique characteristics of airport layouts, surrounding geography, mix of companies and local procedures, etc., there is some generalisability in these findings. Another airport asked for a presentation of LTN's deep dive results, and this was carried out in December 2025 via video call. They found many (though not all) of the issues to be similar to problems they were experiencing, and thought some of the countermeasures would be worth exploring with their business partners. A third major international airport also had similar issues and carried out its own review, and again some of the countermeasures over-lapped. One clear difference was that for this third airport, use of stop bars was considered a more viable solution due to their location on their taxiway system.

On Taxiway Error Reporting

It was noted that five of the events had been classified under the 'Controller Error' rubric rather than 'Taxiway Error', so had not been encoded as taxiway errors in the Safety Dashboard, which limited the use of the Dashboard during the exercise. Additionally, there is usually far more information on causal/contributory factors in incidents happening to inbound compared to outbound aircraft. These two factors blunt the understanding of such events and how to reduce them. Therefore, ways of improving the reporting in both these cases need to be developed and implemented.

Conclusions

A Deep Dive is a very agile process, one that can yield new insights in a very short timeframe. It goes beyond the usual *'what happened and why'* surface layer of incident analysis, to consider issues rooted in the system and conventional ways of working. It considers business choices that can constrain safety, as well as aspects normally considered sacrosanct (e.g. ICAO phraseology), and cultural elements normally kept 'off the table'. The data analytics (presented via a safety dashboard) lent credibility to the process, encouraging the participants to consider 'deeper' factors, and reinforcing the importance of incident reporting. Refinements to data collection at LTN were also identified.

The deep dive process was greatly appreciated by the participants as 'a day well spent', and has resulted in actionable countermeasures, more than half of which have been implemented or are being implemented.

Human Factors often engages with multiple stakeholders when undertaking analysis. The Deep Dive approach would appear to be a worthy addition to the Human Factors practitioner's toolkit. The added value of having a Human Factors professional present is one of clarifying the contributory factors and helping ensure that issues are not muddled or 'lost in translation' during the rich conversational flux of a deep dive exercise. This helps the group steer towards the identification of viable and credible proposals that will effectively counter the underlying contributing factors.

Acknowledgements

The authors wish to thank the pilots of the four airlines (easyJet, Ryanair, Wizzair and Penavia) and additional NATS and LLA personnel for their valuable time and constructive approach during the deep dive.

Funding & Disclaimer

This publication is based on work performed in the HAIKU Project which has received funding from the European Union's Horizon Europe research and innovation program, under Grant Agreement no 101075332. Any dissemination reflects the authors' view only and the European Commission is not responsible for any use that may be made of information it contains.

Reference

Kirwan, B., Bettignies-Thiebaux, B., Cocchioni, M., Baumler, R. and Carrera Arce, M. (2021) Towards a safety learning culture for the shipping industry: a white paper. Malmo: World Maritime University. https://commons.wmu.se/cgi/viewcontent.cgi?article=1000&context=lib_documents

Assessing system safety risks in aircraft landing

Ahmed Jilaau & Gulsum Kubra Kaya

Safety and Accident Investigation Centre, Cranfield University, UK

SUMMARY

Most aviation accidents occur during the landing phase, and accidents are analysed using traditional methods, where individual system components are analysed and, as a solution, the failed component is fixed or removed from the system. This paper aims to analyse the system safety risks in the aircraft landing process by applying the System-Theoretic Process Analysis (STPA) method. STPA enabled a comprehensive analysis of the interactions between system components. The findings revealed 140 unsafe control actions, 142 loss scenarios and 67 safety recommendations for improving the landing process. This study provides an example of STPA application in aviation and valuable insights into accident prevention in the landing phase.

KEYWORDS

System Safety, STPA, Risk Analysis, Aviation, Aircraft Landing

Introduction

Over 63% of aviation accidents occur during the approach and landing phase (Flight Safety Foundation, 2017). The aircraft landing process is complex and requires consideration of the system interactions, including human and software components. However, aviation systems are often assessed using traditional safety assessment methods focusing on individual system components (Bills et al., 2023; Mogles et al., 2018).

Traditional safety assessment methods, such as Failure Mode and Effect Analysis (FMEA) and Fault Tree Analysis (FTA), are linear, focusing on individual system components. Methods like FTA and FMEA are built on the Domino model, where accidents are caused by a chain of events. However, accidents cannot be explained by a static chain of events in complex systems, focusing on system components and aiming to remove failure on the component (Kaya et al., 2021).

System-based methods are developed to address the limitations of those traditional methods. For instance, Leveson & Thomas (2018) introduced System-Theoretic Process Analysis (STPA) for hazard analysis, and Hollnagel (2012) introduced the Functional Resonance Analysis Method (FRAM) to support event analysis and risk assessment. Various research has been conducted using these methods and proved their effectiveness in the analysis (Kaya & Hocaoglu, 2020; Sujan et al., 2024).

As the system-based methods have proven their usefulness and the aviation safety management practices tend to be based on traditional methods, this study aims to revisit the landing accidents by applying STPA to analyse risks involved in the landing phase and provide safety recommendations for aviation safety professionals. The STPA application provides valuable insights into accident prevention in the landing phase.

Method

This study applied STPA to the aircraft landing process of an A320 aircraft in four steps, as described by Leveson & Thomas (2018). These include (1) defining the purpose of the analysis, (2) modelling the control structure, (3) identification of unsafe control actions (UCA), and (4) identification of loss scenarios (LS). This study applies STPA with inputs from authors' industrial and academic expertise, relevant documents (e.g., airline standard operating procedures) and research papers.

In the first step, the purpose of the analysis is identified by defining the system, system boundaries, losses, system-level hazards, and associated system-level constraints. Losses involve something of value that is unacceptable to the stakeholders, such as loss of human life or loss of property, and hazards are conditions in the system that could lead to losses (Leveson & Thomas, 2018).

The next step is to identify the system controllers and their controlled processes to create a hierarchical control structure illustrating the interactions between the controllers. This includes the control actions of the controllers and feedback that controllers receive from the controlled process (Leveson & Thomas, 2018).

In the third step, potential unsafe control actions (UCAs) that could result in hazards are identified and analysed by defining controller constraints (CC) for prevention. A control action can become unsafe either by 'not providing the control action', 'providing the control action', 'providing a safe control action too early, too late, or in the wrong order' or when 'the control action lasts too long or is stopped too soon' (Leveson & Thomas, 2018).

The final step of STPA involves determining loss scenarios by assessing the different causal factors that can lead to the UCAs and their consequences. Based on the loss scenarios, safety recommendations are then proposed for risk mitigation (Leveson & Thomas, 2018).

Results

Define the purpose of the analysis

Four losses (L) were identified in the aircraft landing process: L1- loss of life or injury to people, L2- loss of aircraft or damage to aircraft, L3- loss of or damage to objects outside the aircraft, and L4- financial loss. Next, seven system-level hazards (H) were identified by linking them to these losses and 11 system-level constraints (SC) were linked to each hazard as follows:

- H1: Aircraft deviates from stabilised approach criteria (L1, L2, L3, L4)
  SC1: Aircraft must meet stabilised approach criteria at 1000 ft.
  SC2: If the aircraft deviates from the stabilised approach criteria, a go-around procedure must be carried out.
- H2: Aircraft continues landing without clearance [L1, L2, L3, L4)
  SC3: Aircraft must obtain landing clearance from ATC before commencing landing.
  SC4: If the aircraft continues landing without clearance, ATC must detect the violation, and measures must be taken to prevent any conflicts with traffic.
- H3: Aircraft integrity is lost during landing (L1, L2, L3, L4)
  SC5: Aircraft integrity must be maintained for all conditions of the operation.
  SC6: If aircraft integrity is lost during landing, the fault should be detected, and emergency procedures should be taken to prevent losses.
- H4: Aircraft runway overrun (L1, L2, L3, L4)
  SC7: Brakes and reverse thrust must be applied upon touchdown on the runway.
- H5: Vehicle or aircraft on runway during landing (L1, L2, L3, L4)
  SC8: The runway must be vacated for the landing aircraft.

SC9: If the runway is not vacated, this should be detected, and actions should be taken to prevent collision.

- H6: Communication failure between pilots and ATC (L1, L2, L3)
  SC10: Communication must be maintained between pilots and ATC.
- H7: Inadequate runway surface condition and infrastructure (L1, L2, L3, L4)
  SC11: Runway surface condition must be maintained, and all runway infrastructure, including visual and navigation lights, must be serviceable.

Model the control structure

Figure 1 illustrates the high-level control structure comprising six controllers and 18 control actions. The blue arrows represent the control actions, and the red arrows represent the feedback mechanism between the controller and the controlled process. For example, the airline is responsible for developing and implementing standards and operational procedures for flight operations. The airline, in response, receives feedback via flight crew reports and flight data monitoring programs. Similarly, the flight crew controls the aircraft and monitors the cockpit instruments, while air traffic control (ATC) provides traffic information and landing clearance to the flight crew. In the detailed control structure in Figure 2, 16 controllers and 60 control actions were identified.

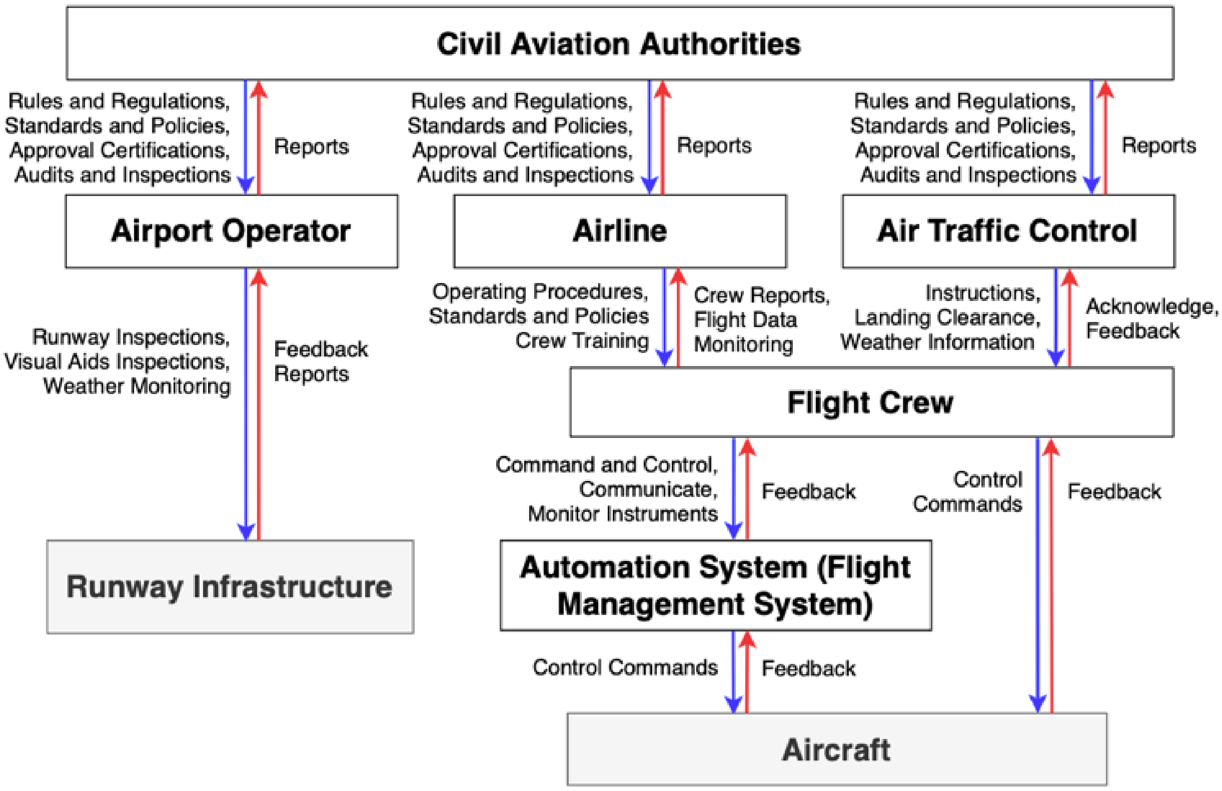


Figure 1: High-level control structure of aircraft landing process

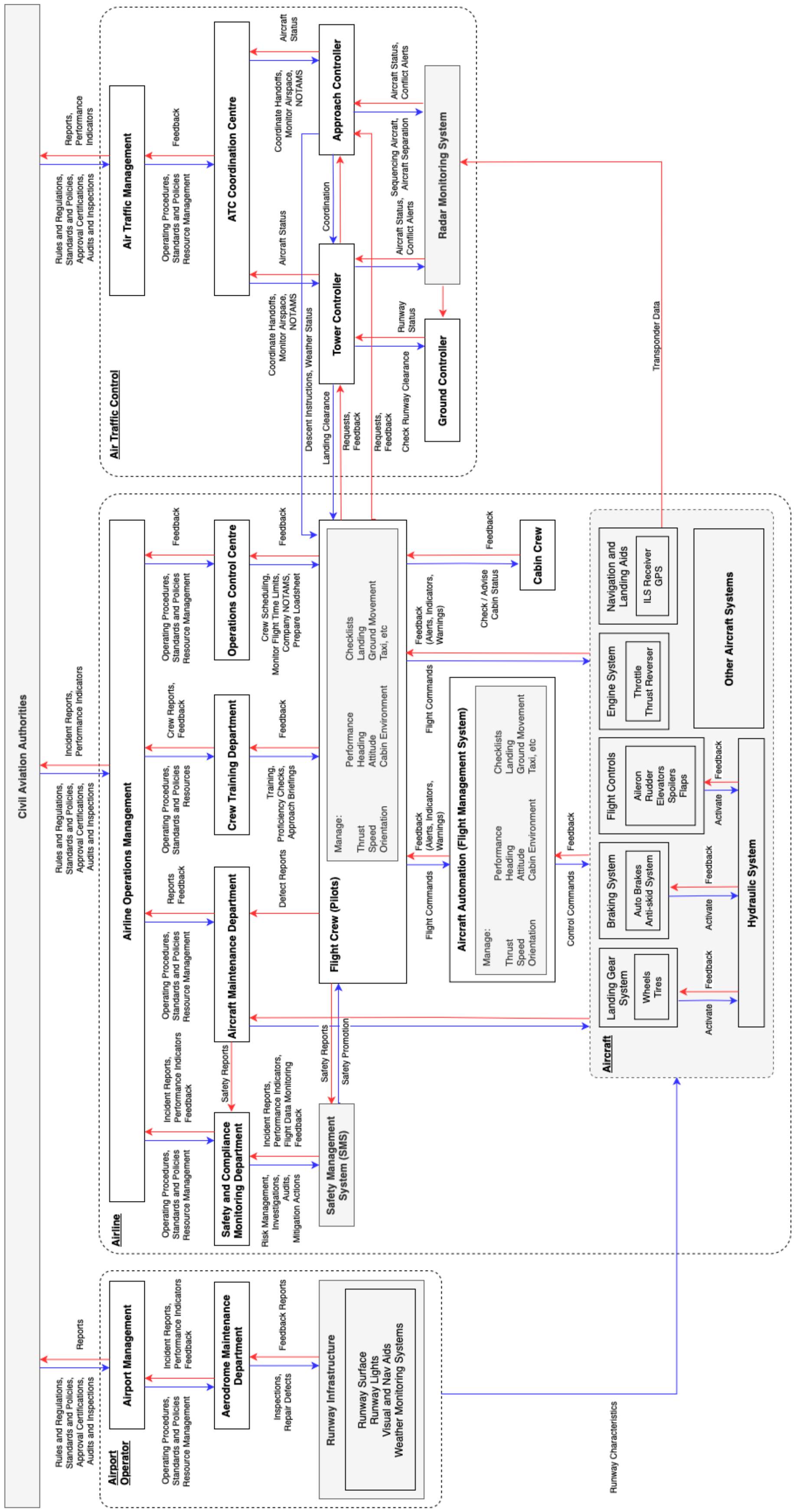


Figure 2: Detailed control structure of the aircraft landing process

Identify the unsafe control actions

In the third step, 140 UCAs were generated considering each control action by considering the four different ways (i.e., not providing the CA, providing the CA, providing CA too early, too late or wrong order, and CA stopped too soon or applied too long) a control action can be hazardous. Table 1 provides a partial list of UCAs for some key controllers.

Table 1: Partial list of UCAs

Controller and Control Actions (CA)	**Unsafe Control Actions (UCA)**
Civil Aviation Authorities: CA1.1: Develop rules and regulations	UCA1.1.1 Rules and regulations not provided by authorities UCA1.1.2 Rules and regulations provided too late by authorities
Airport Operator: CA3.1 Conduct runway inspections	UCA3.1.1 Runway inspections not carried out UCA3.1.2 Runway inspection carried out after landing
Approach Controller: CA6.2 Separation of aircraft in TMA (Terminal Manoeuvring Area)	UCA6.2.1 Separation of aircraft in TMA not provided UCA6.2.2 Separation of aircraft in TMA provided incorrectly UCA6.2.3 Separation of aircraft in TMA not maintained throughout the descent
Tower Controller: CA7.2 Issue landing clearance	UCA7.2.1 Landing clearance not issued to pilots UCA7.2.2 Landing clearance issued to pilots when runway is not clear UCA7.2.3 Landing clearance issued to pilots of wrong flight
Safety and Compliance Monitoring Department: CA10.3: Create safety awareness	UCA10.3.1: Safety awareness not carried out UCA10.3.2: Safety awareness carried out too late UCA10.3.3: Safety awareness continuously not conducted
Aircraft Maintenance Department: C11.1 Carry out maintenance inspections	UCA11.1.1 Maintenance inspections not carried out UCA11.1.2 Maintenance inspections carried out in poor working conditions UCA11.1.3 Defect not detected during maintenance inspections
Crew Training Department: CA12.1 Provide crew training	UCA12.1.1 Training not provided to crew members UCA12.1.2 Inadequate training provided to crew members
Flight Crew: CA14.1 Manage commands on FMS	UCA14.1.1 Command inputs not entered to FMS by pilots UCA14.1.2 Command input entered to FMS incorrectly
FMS: CA16.1 Adjust flight controls	UCA16.1.1 Flight controls not adjusted UCA16.1.2 Flight controls adjusted incorrectly

Identify the loss scenarios

In the fourth step, 142 loss scenarios were identified, explaining the causality of the UCAs. Based on these LSs, 67 different safety recommendations (SR) were identified to mitigate the risks in the aircraft landing process. A partial list of loss scenarios and safety recommendations is described in Table 2.

Table 2: Partial list of loss scenarios and safety recommendations

Unsafe Control Actions (UCA) and Controller Constraints (CC)	**Loss Scenarios (S) and Safety Recommendations (SR)**
UCA1.1.2: Rules and regulations provided too late by authorities (H1, H2, H3, H4, H5, H6, H7) *CC1.1.2: Authorities must provide timely rules and regulations*	S2: UCA1.1.2 might occur when adequate feedback of events not provided to authorities from the operators. This may result in existing regulations to be outdated. *SR2.1: Authorities must review oversight and regulatory audit programs and conduct risk-based audits.*
UCA3.1.1: Runway inspections not carried out (H4, H5, H7). *CC3.1.1: Airport Management must carry out runway inspections.*	S14: UCA3.1.1 might occur due to inadequate procedures related to runway inspections, resulting in FODs on runway during landing. *SR14: Airport Management must develop runway inspection procedures and provide training to staff.*
UCA6.2.2: Separation of aircraft in TMA provided incorrectly (H1, H2, H6). *CC6.2.2: Approach Controller must provide correct separation of aircraft in TMA*	S33: UCA 6.2.2 might occur if the approach controller uses inappropriate radio phraseology when communicating with pilots. *SR33.2: ATC must limit their messages to three topics to avoid confusion.* *SR33.3: ATC must provide one instruction at a time during high workload situation for pilots.*
UCA7.2.2: Landing clearance issued when runway is not clear (H2, H5). *CC7.2.2: Tower Controller must issue landing clearance to pilots when runway is clear.*	S47: UCA7.2.2 might occur if the tower controller had poor situational awareness due to feeling fatigued from increased workload. *SR47: Air Traffic Management should implement a fatigue risk management system for ATC*
UCA10.3.1: Safety awareness not carried out (H1, H2, H3, H4, H5, H6, H7) *CC10.3.1: Safety and Compliance Monitoring Department must carry out safety awareness programs.*	S63: UCA10.3.1 might occur due to inadequate safety culture within the organisation resulting in reduced safety standards. *SR63.1: Airline Management must establish safety policies to promote safety as well as fair and just culture within the organisation.*
UCA11.1.2: Maintenance inspections carried out in poor working conditions (H3). *CC11.1.2: Maintenance Department must carry out maintenance inspections in adequate working conditions*	S69: UCA11.1.2 might occur due to defective lights in maintenance hangar not being rectified, resulting in defects being missed during inspection. *SR69: Maintenance Department must ensure any deficiencies in the facilities are timely corrected to meet the standards as per the regulations and procedures.*
UCA12.1.2: Inadequate training provided to crew members (H1, H2, H3, H4, H6). *CC12.1.2: Training Department must provide adequate trainings to crew members.*	S76: UCA12.1.2 might occur due to inadequate crew training syllabus. *SR76: Crew Training Department must implement Evidence Based Training (EBT) and assess crew competencies.*
UCA14.1.2: Command input entered to FMS incorrectly (H1, H2, H3, H4). *CC14.1.2: Pilots must have correct inputs to the FMS.*	S89: UCA14.1.2 might occur if the pilot workload was high and situational awareness was low *SR89.1: Pilots must ensure there is adequate crew resource management between the crew members.*

	SR89.2: Airline Management must promote importance of performing go-arounds if the approach is unstable
UCA16.1.1: Flight controls not adjusted (H1, H3, H4). *CC16.1.1: Autopilot must adjust flight controls.*	S131: UCA16.1.1 might occur due to erroneous data from sensors on aircraft. *SR131.2: Pilots must regularly monitor / detect errors.*

Discussion

The STPA application enabled a comprehensive assessment of the aircraft landing process. The study provided 67 safety recommendations for improving safety in addition to the system level and controller constraints identified.

The findings showed that most safety recommendations were related to organisational factors, such as training, operational procedures, safety management, crew scheduling and communication, which could be due to the method as organisational factors tend to be identified more using STPA (Kaya et al., 2021). For instance, poor decision-making due to inadequate training or lack of knowledge was found to be a critical causal factor for many UCAs. With over more than 50% of aircraft accidents linked to poor decision-making by flight crew (Harris & Li, 2017), implementation of Evidence-Based Training (EBT) program for flight crew based on the guidance provided by ICAO (2013) is proposed as a key safety recommendation (SR76). This competency-based approach comprises eight main flight crew competencies, including the application of procedures, communication, aircraft flight path management– automation, aircraft flight path management – manual control, leadership and teamwork, problem-solving and decision-making, situational awareness and workload management. IATA (2024) recommends the 'application of knowledge' as an additional competence.

By considering the most likely threats to flight operation based on historical data, EBT guides airline operators to ensure the flight crew are competent to operate the flights safely. Unlike the traditional task-oriented approach of crew training, EBT promotes effective management of unexpected situations, as the lessons learnt can be utilised in various scenarios instead of pre-defined ones (IATA, 2024).

In addition, many loss scenarios, such as S89 and S92, are contributed by poor communication and ineffective task management. Based on research done by NASA, such critical factors in the cockpit account for most aviation incidents and accidents caused by human error, compared to technical issues of operating in a cockpit (Shappell et al., 2006). However, human error is a consequence, not a cause. Thus, research focused on systems to enable humans to do the right thing every time. This is also further supported by the fact that 50% of accidents and investigation reports by NTSB had mentioned Crew Resource Management (CRM) as a contributory factor, with 71% of accidents occurring during the landing phase of the flight (Wagener & Ison, 2014). As a result, CRM training for crew members has become a mandatory requirement for all commercial airlines and is a critical component of their non-technical training (Harris et al., 2024). The approach towards CRM has also evolved over the years with an additional focus on proactively identifying errors and mitigating their consequences, known as error management (Hayward & Lowe, 2017). With CRM found to be effective in improving the decision-making, situational awareness, communication, and leadership of flight crew (Salas et al., 2001), it is highly recommended that pilots ensure CRM in the cockpit is managed efficiently (SR89.1).

Furthermore, the STPA application also provided recommendations for managing unstable approaches, which were identified as a major contributing factor in 14% of approach and landing incidents, including runway excursions (IATA, 2022). One such recommendation is for airlines to implement and promote flight crew awareness of the importance of initiating a go-around when the approach becomes unstable (SR89.2). This has also been supported by a study by the Flight Safety Foundation (2017), which showed that a significant majority (83%) of runway excursions and over half of all accidents could have been avoided by performing a go-around. While there are multiple reasons why the compliance rate with the go-around policy during an unstable approach is low (3%), airlines must take measures to improve safety culture and awareness (SR63.1) among flight crew by demonstrating that the airline management will support the flight crew judgement to go-around in such scenarios without any penalty for operational disruption. Moreover, the stable approach criteria defined in the operational procedures must be regularly reviewed and updated to remove any subjectivity in their decision. The regulatory authorities must also be actively involved in improving the compliance rate for the go-around policy by reviewing the existing oversight audit programs (SR2.1) to check for go-around compliance among operators (Flight Safety Foundation, 2017).

Another key finding was the unsafe conditions created due to inadequate communication between the flight crew and ATC, as described in S33. One of the contributing factors to such misunderstanding is improper usage of radio phraseology, its speed, and language barriers, especially for non-native English speakers (EUROCONTROL, 2017). Therefore, it is essential for controllers to use the correct radiotelephony procedures as recommended by ICAO Doc444 and Doc9432 (International Civil Aviation Organization, 2007a, 2007b). Moreover, as recommended in SR33.2 and SR33.3, controllers should limit their instructions to the flight crew to three topics to avoid any misunderstanding. They must provide one instruction at a time for non-native English speakers during approach and landing where the workload is high (Barshi & Farris, 2013).

With the increase in air traffic means, air traffic control officers (ATCO) are subjected to high workload situations while managing multiple flights in airspace, resulting in scenarios such as S47. Therefore, it is highly recommended to establish an effective fatigue risk management system (FRMS) for ATCOs (SR47) in accordance with ICAO fatigue risk management guidelines to ensure adequate rest periods, rostering, and rest facilities are provided by the operator. The rostering of ATCOs should use the principles of fatigue science (International Civil Aviation Organization, 2016). A study by Li et al. (2020) found that high traffic volumes contribute to the increased mental workload of ATCOs, which contributes to more airspace incidents; hence, FRMS should focus on providing adequate breaks to maintain their cognitive resources.

While this study offers valuable insights into managing risks in the landing phase, the study has limitations. The STPA analysis could have been conducted with the involvement of subject matter experts (SMEs). This study is done as a desk study rather than involving SMEs. While this might limit the analysis, the authors used their academic and professional experience in safety science and aviation safety management.

Conclusion

STPA was applied to the aircraft landing process to examine the interactions between the various stakeholders involved and identify its system safety risks. Each controller is responsible for critical control actions, and a deficiency in a single control action could contribute to unsafe conditions. The STPA analysis resulted in a comprehensive risk analysis with the generation of many loss scenarios and safety recommendations focused on improving the system, including human and organisational factors.

References

Barshi, I., & Farris, C. (2013). Misunderstandings in ATC communication: language, cognition, and experimental methodology. In *Misunderstandings in ATC Communication: Language, Cognition, and Experimental Methodology*. Ashgate Pub. Co. https://doi.org/10.1080/00140139.2013.868633

EUROCONTROL. (2017). *European Action Plan for the Prevention of Runway Incursions (v3.0)*. https://www.eurocontrol.int/sites/default/files/2019-06/european-action-plan-prevention-runway-incursions-v3.pdf

Flight Safety Foundation. (2017). *Go-Around Decision-Making and Execution Project*. https://flightsafety.org/toolkits-resources/go-around-project-final-report/

Harris, D., Chan, W. T. K., Chatzi, A., Griebel, H., Li, W. C., Lu, T. T., McCarthy, P., Nakanishi, M., Plioutsias, T., & Ziakkas, D. (2024). Report of the working group to identify future challenges faced by the implementation of resource management in remote and distributed teams. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *14692 LNAI*, 190–200. https://doi.org/10.1007/978-3-031-60728-8_15

Harris, D., & Li, W. C. (2017). Decision making in aviation. *Decision Making in Aviation*, 1–487. https://doi.org/10.4324/9781315095080/DECISION-MAKING-AVIATION-HARRIS/ACCESSIBILITY-INFORMATION

Hayward, B. J., & Lowe, Andrew. (2017). *Safety and error management: The role of crew resource management*. 107–119. https://doi.org/10.4324/9781315181837-11

ICAO. (2013). *DOC 9995, Manual of Evidence-based Training* (1st ed.). International Civil Aviation Organization.

International Air Transport Association. (2022). *Examining Unstable Approaches-Risk Mitigating Efforts*.

International Air Transport Association. (2024). *Evidence-Based Training Implementation Guide* (2nd ed.). International Air Transport Association. https://www.iata.org/contentassets/632cceb91d1f41d18cec52e375f38e73/ebt-implementation-guide.pdf

International Civil Aviation Organization. (2007a). *Doc 4444: Procedures for Air Navigation Services - Air Traffic Management* (15th ed.). International Civil Aviation Organization.

International Civil Aviation Organization. (2007b). *Doc 9432: Manual of Radiotelephony* (4th ed.). International Civil Aviation Organization.

International Civil Aviation Organization. (2016). *Fatigue Management Guide for Air Traffic Service Providers* (1st ed.). International Civil Aviation Organization.

Leveson, N. G., & Thomas, J. P. (2018). STPA Handbook. In *2018*.

Li, W. C., Kearney, P., Zhang, J., Hsu, Y. L., & Braithwaite, G. (2020). The analysis of occurrences associated with air traffic volume and air traffic controllers' alertness for fatigue risk management. *Risk Analysis*, *41*(6), 1004–1018. https://doi.org/10.1111/RISA.13594

Salas, E., Burke, C. S., Bowers, C. A., & Wilson, K. A. (2001). Team Training in the Skies: Does Crew Resource Management (CRM) Training Work? *Https://Doi.Org/10.1518/001872001775870386*, *43*(4), 641–674. https://doi.org/10.1518/001872001775870386

Shappell, S., Detwiler, C., Holcomb, K., Hackworth, C., Boquet, A., & Wiegmann, D. (2006). Human Error and Commercial Aviation Accidents: A Comprehensive, Fine-Grained Analysis Using HFACS. *Publications*. https://commons.erau.edu/publication/1218

Wagener, F., & Ison, D. C. (2014). Crew Resource Management Application in Commercial Aviation. *Journal of Aviation Technology and Engineering*, *3*(2), 02. https://doi.org/10.7771/2159-6670.1077

Augmenting aviation incident analysis with Artificial Intelligence, and the curse of dimensionality

Kirwan, B.,[1], Elliott, R.[2] Bolger, L.,[2] Biliri, E.,[3] Koussouris, S.,[3] Durante, N.G.,[4] Wright, P.,[4] Newman, D.,[5] and Bettignies-Thiebaux, B.[1]

[1]EUROCONTROL, [2]London Luton Airport, [3]Suite5, [4]ENG, [5]NATS

SUMMARY

The HAIKU project aims to explore Human-AI Teaming via six aviation use cases. One of these focuses on the reduction of airside incidents at a UK international airport via AI-based analysis of safety events and occurrences at this airport. In particular, three incident types were selected: Hold-Point Busts, where an aircraft is instructed to wait at a Hold-Point on the taxiway system but doesn't; Pushback Errors, wherein an aircraft is incorrectly pushed back from the stand so that it is facing the wrong direction; and Taxiway Errors, wherein the flight crew of an aircraft make an error on their routeing along the taxiway system.

A seven-year incident dataset was processed by two AI companies, with input from the airside operations team and the local air traffic control organisation to help understand the data and various factors affecting safety. Despite more than half a million aircraft movements during this period, the corresponding number of incidents (for the three incident types of concern, numbering hundreds in the same seven-year analysis period) proved problematic for AI developers. This was in part due to what is known as 'the curse of dimensionality', wherein there are too many dimensions (characteristics related to incident causation) given the amount of data. Nevertheless, a Dashboard based on the data analytics, together with airport operational expertise, led to new, actionable insights that have enhanced safety for two of the incident categories. This paper presents the Dashboard and its usage in incident reduction at the airport.

KEYWORDS

Aviation, AI, Incident Analysis, Safety Dashboards, Human Factors

Background

London Luton Airport (LTN) is the fifth busiest airport in the UK, operating a single runway for commercial and business operations, and is a hub for low-cost airlines including EasyJet, TUI, Wizzair and Ryanair, alongside a number of business jet operators and ground handling services (Figure 1), with air traffic services provided by National Air Traffic Services (NATS).

Figure 1. Illustration of LTN Partners and Ground Handling

Safety performance at LTN is excellent, but there are three incident types that have proven hard to eradicate (see Figure 2):

1. **Hold-Point Bust (HPB)**: wherein an aircraft is instructed to proceed to a hold-point on the taxiway system and await further instructions, but the aircraft continues through the hold-point. This has the potential to result in a taxiway collision.
2. **Taxiway Error**: wherein an aircraft turns the wrong way on the airport taxiway system (or continues when it should turn or wait), which can also result in a collision or (far more likely) in an aircraft having to take a circuitous route to where it should be, resulting in delays on the ground if the airport is busy.
3. **Pushback Error**: for example, when an aircraft is pushed back by a tug from its gate but is reversed to the left instead of the right, so that it ends up facing the wrong way on the taxiway system or terminates in the incorrect geographical position.

Figure 2. Three incident types (from left to right) Hold-Point Bust, Taxiway Error, Pushback Error

These incident types may be considered precursors to far more serious incidents and accidents, e.g. two aircraft colliding on the taxiway or apron area, or even on the runway. One of the prime safety goals at any airport is to reduce the occurrence rate of such incidents to a minimum, ideally eradicating them entirely. However, in practice this has proven difficult as there are many factors at play, and airports tend to be complex and dynamic systems-of-systems involving the interplay of multiple airlines and business jet users, air traffic control, and a host of ground handling services (fuelling, baggage handling, aircraft marshalling, etc.). Although LTN has an excellent safety record, these three incident types continue to occur each year. In an effort to raise their level of safety, LTN organisations enlisted the aid of the European research project HAIKU (https://haikuproject.eu), which aims to explore Human-AI Teaming via a set of aviation use cases. Two AI companies, plus the local LTN Air Traffic Control Tower (run by NATS), joined the airport authority (London Luton Airport or LLA – the overall airport operator) and EUROCONTROL to see if AI could help identify new ways to reduce the occurrence of these incidents. The support to LTN occurred in two main steps: firstly, collecting and processing the data from LTN into an AI model to help identify incident factor relationships, and secondly visualising the incident data and associated factors in a Safety Dashboard.

Collecting and Processing the Data

Firstly, data collection is performed, gathering relevant information from various sources. Next, data preprocessing takes place, involving cleaning the dataset by handling missing values, removing duplicates, and standardising formats and content, since most of the inputs are given in the text format (i.e., manually entered from pilots/operators). Then, Exploratory Data Analysis (EDA) is conducted to understand patterns, distributions, and correlations through visualisations and statistical techniques. Feature engineering follows, where meaningful features are selected, transformed, or created to improve model performance. These steps are illustrated in the Figure 3.

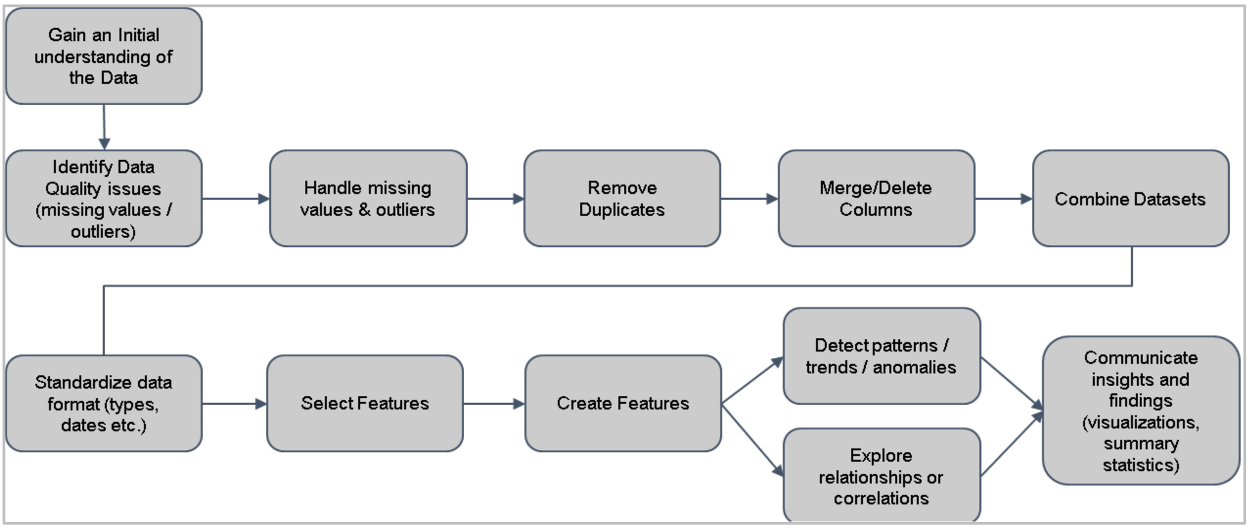


Figure 3. Steps in Data Collection, Preparation and Model Building

Additional information is then added from external sources, such as weather conditions, number of sectors the flight had flown, etc. Additional data inclusion is dependent on input from LTN business partners as to what they judge to be important. This is where visits by the date scientists to the airport and tower and the stakeholder meetings pay dividends in gaining an accurate model.

The initial hope had been that an AI model could be developed based on the use of data science techniques (machine learning applied to large datasets), which would pinpoint coincidental factors that led to these incidents. However, after some six months of data preparation and analysis, it became apparent that there were insufficient numbers of incidents given the diversity of unique factors or factor combinations in individual incidents in the dataset. This is a well-known phenomenon in data science, called the 'curse of dimensionality'[1]. Ironically, if there were many more incidents, the likelihood of developing a good model to help reduce them would increase.

The analysis also revealed contributory factors that could not be deduced from the information in the datasets, as well as cases where the incident cause would not be possible to be predicted beforehand (e.g. distractions in cockpit), making it hard to derive a predictive model. Despite such limitations, the project continued into the development of incident visualisation approaches in the form of a four-tier Safety Dashboard.

Safety Dashboard

The Airport Safety Watch (ASW) Dashboard (henceforth called simply 'the Dashboard) was developed by the data science teams in conjunction with key LTN stakeholders (in particular LLA and NATS). It has four tabs or layers, each one offering different insights into the incidents and the circumstances under which they happened:

1. The **Past Overview** tab allows users to see aggregate information for incidents & flights within a specific time period. Various dimensions (temporal, traffic-related, airline-related, ongoing construction work, etc.) are explored to see if they affected incident occurrence.
2. The **Comparison** tab allows users to review incident information for a specific time period compared to another, e.g. comparing June 2024 against June 2023. All airports make these kinds of comparisons since weather and travel patterns are both very seasonal in nature. The same dimensions as in the *Past Overview* tab are explored to see how they affect incident occurrence and are compared across the two periods.
3. The **Zoom-In** tab allows users to:
   i. Create custom incident subsets using an adaptable tree view and explore aggregated & low-level (individual incident level) information from incident reports;
   ii. Zoom into an incident, review its detailed information and explore airport congestion insights around the time it happened.
4. The **Monitoring** tab offers real-time information regarding incidents and airport conditions, including weather.

The four tabs (pages) offer customization options that enable users to gain quick insights and 'drill down' to specific parts of the data, depending on user needs. As an example, metrics can be shown in the form of absolute numbers (e.g. number of incidents) or normalized across dimensions of interest (e.g. number of incidents across number of flights). Users can also customize whether the visualisations show aggregate information about all three incident types or split per type. Figure 4 shows (upper part on left) an extract from the *Past Overview* mode, highlighting 'hotspot' areas on the airport surface layout (stands, apron, hold-points, taxiways and construction areas).

[1] The curse generally refers to issues that arise when the number of datapoints is small relative to the intrinsic dimension of the data. For LTN incidents, of which there are several hundred, there are >20 aspects (dimensions) of interest (time of day, airline, congestion etc.). Each incident must have all these aspects described, and there need to be many more incidents for effective model-building. See: https://en.wikipedia.org/wiki/Curse_of_dimensionality

Figure 4. Extract from Airport Safety Dashboard – Past Overview mode

Figure 4 also shows (lower part, left and centre) a histogram of event occurrence according to time (selectable – a day, a month, a year etc.) and traffic levels, to see for example if traffic congestion was a factor. It is also possible to look at time of day, how many sectors the crew of the aircraft involved had flown that day, weather aspects, ongoing construction affecting stands or taxiways, direction of runway usage at the time, whether the aircraft had suffered or was incurring a delay, etc. Eliciting and visualising these factors required numerous meetings between the data scientists and LTN personnel, including several visits to the airport and the control tower.

Additional information is provided within charts either by 'hovering' with the mouse or trackpad over various elements, or through certain selections and other interactions. For example, Figure 5 shows two different layers of the incidents map (compared to the default 'heatmap' view shown in the Past Overview in Figure 4), the left view showing exact incidents' stands and hold-points, the right view showing additional information provided to the user for the incidents that happened at a specific stand/hold-point. LTN airport business partners were particularly interested in these views.

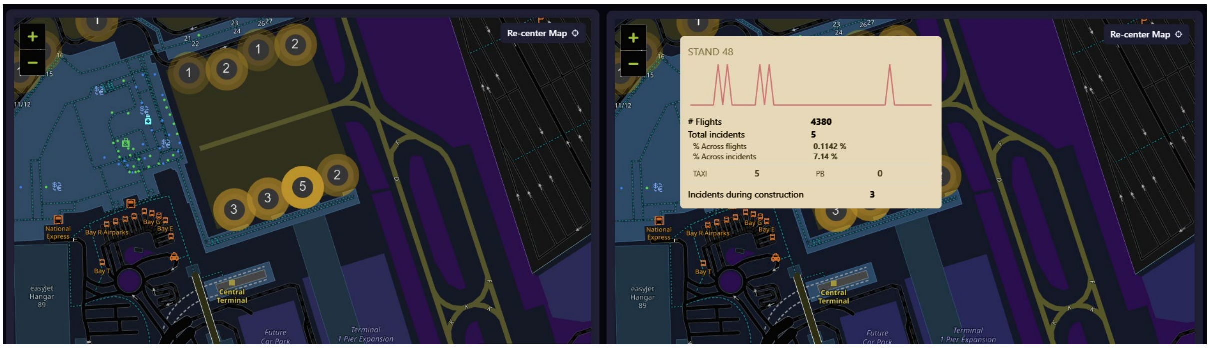


Figure 5. Extract from Airport Safety Dashboard – Airport Map

The Monitoring view of the dashboard in Figure 6-7 provides a comprehensive visualization of flight operations and error statistics, enabling users to analyse trends and identify potential issues efficiently. The Flight Data section presents an interactive timeline of flight distribution, allowing users to zoom into specific periods for detailed analysis, highlighting peak operational hours and key statistics, including runway-specific activity. The Errors section tracks runway changes over time and categorizes operational disruptions (pushback, taxi, and holding point errors), while also integrating maintenance work data. Interactive elements reveal error durations, daily and hourly distributions (normalized by flight volume), and errors associated with specific stands. Additionally, a spatial map visualizes errors linked to stand positions, supporting in-depth situational awareness.

Figure 6. Extract from Airport Safety Dashboard – Monitoring Mode

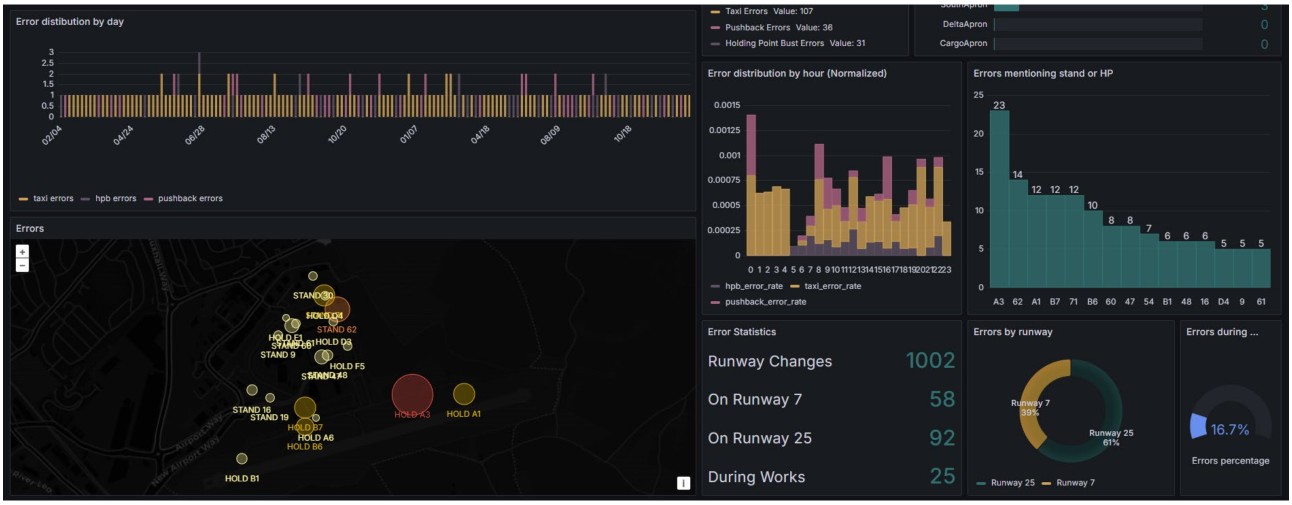


Figure 7. Extract from Airport Safety Dashboard – Monitoring Mode: Focus on errors

The Zoom-in mode of the dashboard allows users to query the unstructured and semi-structured information found in the incident reports and the questionnaires filled in by the involved stakeholders (crew members) after the incident. As an example, insights into the familiarity of the captain (or first officer in case he/she was in control at the time of the incident) with the LTN layout, the perceived ability or confidence to challenge ATC instructions, etc., can be very useful for identifying causes or contributory factors of incidents and either further grouping some incidents according to such latent factors or diving into the details of specific incidents of interest. Dashboard users can browse through answers to such questions and also use them as filters to examine custom incident subsets such as “the incidents for which the involved parties stated that the instructions were clear, that they are familiar with LTN layout and that the conditions did not change in the time leading to the incident” and check for patterns regarding their time or location.

The analysis work is continuing, and ground collisions are to be added to the Dashboard in 2025. The Dashboard is now updated on a weekly basis and is used by LTN operational safety staff to talk to ground handlers and pilots about safety risks. The Dashboard lends credibility to such safety messages, as the safety reps can show the statistics and trends and highlight what matters and what does not appear to matter. The Dashboard is therefore feeding and supporting safety conversations. This can also inspire more people to report incidents fully in the first place and be prepared to add more detail – particularly on contributory factors such as fatigue, loss of situation awareness, expectation bias etc. – that might prove more critical in AI model development.

Safety Outcomes from Use of the Dashboard

A safety meeting of around 25 key airport business partners in July 2023, where the Safety Dashboard was presented and discussed between the business partners and the data scientists, proved particularly constructive and generative, leading to the identification of several new safety insights, two of which are described below.

- Stands 62 and 71 were highlighted by the data analysis presentation as being more prone to taxiway error. Partners noted that Stand 62 has no sign, which might contribute to error rates; it is for business jets rather than commercial jets, and many business jet pilots are unfamiliar with LTN's layout etc. Stand 71 is a cul-de-sac and not be as well signposted as other stands.
- A particular hotspot on the taxiway system was identified. Changes were subsequently made to the airfield, including directional paint markings, new Delta/Foxtrot signage (see Figure 8) on taxiway Alpha. NATS have also begun more 'defensive controlling' in relation to the Alpha/Delta/Foxtrot intersection. There have been two incorrect taxi events in the eight months since the new signage has been installed, compared to six events in the previous eight months.

Figure 8 – Signage change at taxiway intersection

The dashboard was also very helpful to the Stack in ruling out potential factors, for example many business partners thought that these incidents occurred at busier times, however the dashboard showed that it was during the quieter times (Figure 9). Similar results were obtained when looking at other 'usual suspects' for incident causation, such as weather (in particular, low visibility), also found not to be a strong factor.

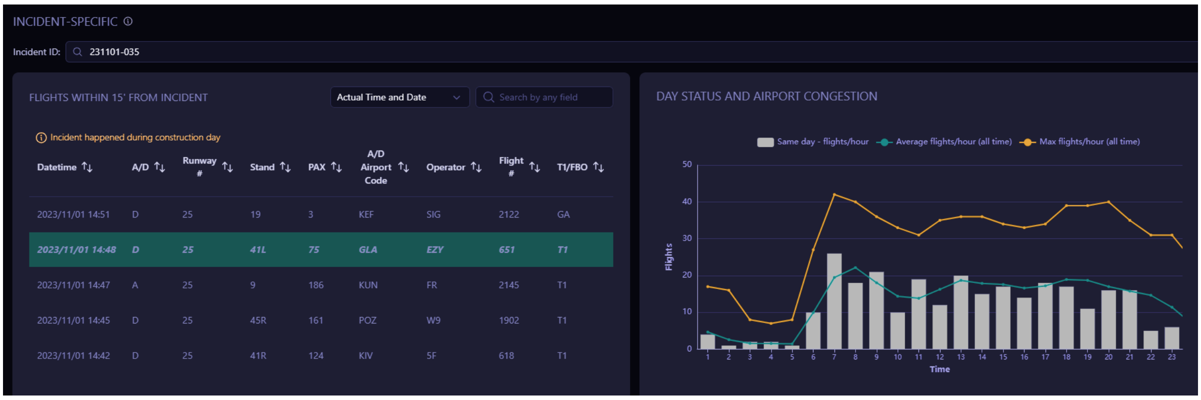


Figure 9: Zooming in on flight congestion as a factor for particular incidents

Such insights as the above were the result of a partnership between the Stack stakeholders who had significant operational expertise, and the data scientists working with the data and presenting and exploring the various views of the Dashboard. In essence this was an early example of Human-AI Teaming, one in which both parties were required to achieve fresh safety insights. The Dashboard alone did not identify the insights, and the LTN partners could only identify such insights when they had a richer visualization of the operational incident 'landscape', as well as ways to interrogate and cross-examine that landscape.

Conclusion

This early attempt at using AI for safety highlights the difference between the general AI hype and what is realistically possible. Yet even this level of data, scientific analysis has augmented the airport's ability to manage safety. Despite the dimensionality issue, the Dashboard has helped LTN identify specific hotspots and key stands and taxiway intersections where more incidents occur, and resultant measures (signage, ways of working and education) are having a positive impact on reducing incident rates. As well as this direct impact, the Dashboard also helps Safety Stack partners understand which factors are *not* key, so that they can avoid spending resources on areas unlikely to have a material effect on incident rates.

Future work will integrate vehicle-vehicle and vehicle-aircraft collision incident data into the Dashboard. As incident data accrues, it is hoped one day that a true predictive AI model can be constructed. In the meantime, data science partnered with human expertise will continue to advance safety learning at LTN.

Funding

This publication is based on work performed in the HAIKU Project which has received funding from the European Union's Horizon Europe research and innovation program, under Grant Agreement no 101075332. Any dissemination reflects the authors' view only and the European Commission is not responsible for any use that may be made of information it contains.

Can AI Recognise Pilot's Vocal Emotional Expression under Emergency Situations?

Wen-Chin Li, Kuang-Lin Hsieh, Jeremia Pramudya & Declan Saunders

Safety and Accident Investigation Centre, Cranfield University

SUMMARY

The development of Artificial Intelligence (AI) in the field of voice recognition has prompted interest in the field of emotional voice recognition (EVR); EVR is now one of the key challenges in the applications of Natural Language Processing (NLP). When conducting an accident investigation, the voice data from the cockpit voice recorder (CVR) usually provide significant evidence of the pilot's mental state, which can support some hypotheses on occurrences by investigators, especially the verbal speech between pilots and air traffic controllers. In the past, emotion analysis mainly relies on image and text analysis technology. With the development of large language models (LLM) and AI; such as ChatGPT, Llama, and Perplexity, EVR has become possible. This research aims to explore the potential of using a Recurrent Neural Network (RNN) and the open-source dataset, Toronto Emotional Speech Set (TESS), to identify the pilot's speech and emotions in emergencies. Further research may combine voice with physiological data, and with facial expressions to serve the purposes of operational and safety monitoring.

KEYWORDS

Artificial intelligence, Deep Learning, Emotional Voice Recognition, Large Language Model, Machine Learning

Introduction

The human ability to perceive nuances in verbal communication is extremely sophisticated; with accuracy and capacity achieved over years of evolution to communicate consciousness and perception (Alharbi et al., 2021). Recent studies into EVR, using artificial intelligence (AI), have demonstrated a capacity for feature extraction, such as pitch, tone, and speech rate. These features correlate with human emotion: a high pitch and speech rate could indicate excitement, while a slower speech rate and tone could indicate sadness (Cowen & Keltner, 2017). In aviation, pilot decision-making during emergencies is crucial to ensuring the safety of passengers and crew members. When faced with unexpected events, pilots must quickly respond and make decisions based on limited information and high-pressure conditions. These situations reduce the pilot's ability to fully utilise the resources and information available, increasing the chance of human error. Emotional responses, such as anxiety, stress, or fear, can impair a pilot's attention, memory, and judgment, thus affecting their choices during critical moments. Emotional interference with cognitive functions may lead to delays in decision-making or poor judgment, which could severely threaten flight safety (Nocak & Mrazova, 2015).

This research aims to use Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) models to recognise the emotion in verbal communications in high-stress environments. Deep Learning (DL) has emerged as a key research area in Machine Learning (ML), gaining attention for its ability to process raw data without manual feature extraction. In recognising data, DL surpasses traditional methods by automatically detecting complex patterns and handling unlabeled data. DL has made significant advancements in natural language processing (NLP), achieving breakthroughs in applications like speech recognition and other language-related tasks (Barhoumi & BenAyed, 2025). A Recurrent Neural Network (RNN) is a type of artificial neural network with directed cycles, allowing information to persist. Unlike feedforward networks, RNNs retain past inputs, making each output dependent on both current and previous inputs (Sherstinsky, 2020). Consider a general recurrent network with input $\{x^{(t)}\}$, output $\{z^{(t)}\}$, and target sequence $\{y^{(t)}\}$, where $z^{(t)} \in \mathbb{R}^p$ and $y^{(t)}a \in \mathbb{R}^p$. The data-generating process (DGP) is modeled as:

$$y^{(t)} = z^{(t)} + \varepsilon^{(t)}, \text{ for } t \in \mathbb{Z},$$

Where $\{\varepsilon^{(t)}\}$ is a sequence of independent and identically distributed (*i.i.d.*) random vectors. This additive error assumption is widely employed in statistical modeling and aligns with commonly used loss functions designed to quantify the discrepancy between $y^{(t)}$ and $z^{(t)}$ (Zhao et al., 2020). Python has become a dominant programming language in EVR due to its extensive libraries and frameworks that facilitate ML and DL applications. The librosa library is widely used for feature extraction from audio signals, enabling researchers to analyse MEL-frequency cepstral coefficients (MFCCs), pitch, and energy levels essential for emotion detection (Barhoumi & BenAyed, 2025). DL frameworks like TensorFlow and PyTorch provide robust tools for building RNN, LSTM, and gated recurrent unit (GRU) models, which are effective in processing sequential voice data (Abdel-Hamid et al., 2014). These models learn temporal dependencies in speech, allowing the system to recognise emotions from voice modulation (Luo et al., 2017).

Methods

3.1 Data collection and loading:

This research utilises the open-source database, TESS (Toronto Emotional Speech Set). It is then randomly divided into two groups (80% of training data and 20% of testing data) to build the accuracy of the model TESS (Toronto Emotional Speech Set):

TESS is a speech dataset that contains speech samples from two female actresses, aged 29 and 64, using a set of 200 target words expressing seven different emotions: anger, disgust, fear, happiness, pleasant surprise, sadness, and neutral. The dataset includes 2,800 samples in total.

3.2 Feature Extraction:

Mel Frequency Cepstral Coefficients (MFCC): Designed on the Mel frequency scale, MFCC attempts to simulate how the human ear perceives and processes speech, making it one of the most commonly used feature extraction methods in speech processing (Barhoumi & BenAyed, 2025). The MFCC will extract audio from a WAV file format, and apply coefficients in a 13-dimensional space, one dimension being the loudness and frequency of the voice (Ittichaichareon et al., 2012). Figure 2 shows the MFCC creating a vector consisting of loudness, pitch, frequency, and speech sounds including tone, rhythm, and articulation respectively.

3.3 Training Model with neural network:

Features extracted will be fed into the RNN and LSTM network to capture the temporal dependencies inherent in speech signals. LSTM is a specialised type of RNN designed to overcome the vanishing gradient problem that happens in RNN. LSTM contains a memory cell that can store information over long periods and has several gates (input, forget, and output gates) that control the flow of information in and out of the cell. These gates enable the LSTM to decide which information to retain, which to forget, and which to output. This allows LSTM to capture long-term dependencies in sequences. LSTM performs better than

RNN in tasks requiring the model to retain information for longer durations, as they can maintain and modify the internal state based on both the current input and previous context.

3.4 Applying the audio into the system:

This research aims to analyse the emotion with the system trained via DL, thus, the audio of the communication between the pilot of Flight 1549 and ATCOs for 90 seconds released by the Federal Aviation Authority (FAA) is applied. The proposed methodology for the EVR analysis consists of the following steps seen in Figure 1 below.

Figure 1 Flowchart of the overall architecture

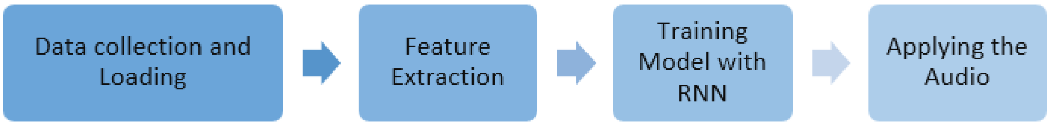


Figure 2. An example of the anger extracted audio

```
Sample Rate: 48000 Hz, Signal Length: 166567 samples
MFCC Features: [ 1.04644124 -6.68054316 -0.48840813 -0.60830217 -0.27202627  0.30199612
 -0.94593789  0.12241271 -0.50382968 -0.02061761 -0.28371644 -0.29715581
  0.39338679]
```

Results

In this study, we employ Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) for emotion recognition in cockpit dialogues. The model achieved an accuracy of 89.54% on the training data and 85.26% on the test data, shown in Table 1, demonstrating strong performance in identifying emotions. The pilot had 8 speeches to communicate with ATCO in 90 seconds, as shown in Table 2. The result demonstrated that AI recognised the pilot's emotion of fear (97.0%) at the beginning of the bird strike, followed by anger (2.8%), and finally disgust and pleasant surprise (0.1%). In those 90 seconds of communication, the emotion was also detected as sadness (51.1%), fear (47.7%) and neutral (1.2%).

Table 1: The accuracy of the model

Data (Source: TESS)	Method	Feature Extraction	Accuracy
Training Data	RNN + LSTM	MFCC	89.54%
Test Data			85.26%

Table 2: The emotion of the pilot predicted from the model

audio \ emotions	angry	fear	disgust	happy	pleasant surprise	sad	neutral
Pilot's 1st speech: 00:00 – 00:06	2.8%	97.0%	0.1%	0.0%	0.1%	0.0%	0.0%
Pilot's 2nd speech 00:10 – 00:10	0.7%	85.8%	0.0%	0.0%	0.0%	2.4%	11.1%
Pilot's 3rd speech 00:33 – 00:35	7.6%	88.9%	0.0%	0.5%	0.0%	0.5%	2.5%
Pilot's 4th speech 00:39 – 00:39	0.1%	70.4%	0.1%	0.0%	0.0%	27.7%	1.7%
Pilot's 5th speech 00:48 – 00:52	0.2%	99.8%	0.0%	0.0%	0.0%	0.0%	0.0%
Pilot's 6th speech 00:58 – 00:58	0.1%	42.1%	0.0%	0.0%	0.0%	54.0%	3.8%
Pilot's 7th speech 01:20 – 01:21	51.3%	31.7%	5.1%	4.8%	0.0%	7.1%	0.0%
Pilot's 8th speech 01:23 – 01:24	0.0%	47.7%	0.0%	0.0%	0.0%	51.1%	1.2%

The study also conducted human evaluation on the same 90 seconds of audio transcript, to evaluate and compare the emotional recognition conducted by the AI. The evaluation was conducted as either an emotion was present or was not present, instead of a percentage of emotion in each speech. Based on an analysis with 34 subject matter experts (SME), it was concluded that the highest emotion detected in each transcript was (1) fear (8 out of 8 speeches), followed by (2) neutral (8 out of 8 speeches), then (3) anger (5 out of 8 speeches) and finally (4) disgust (4 out of 8 speeches). The emotional recognition of the human analysis revealed a significant deviation from the AI model, which demonstrates the complexity and diverse nature of EVR.

Discussion

This research successfully applied EVR modeling to analyse the cockpit dialogue of pilots and ATCOs in emergencies, effectively identifying various emotional states. Table 2 contains the weighting of each emotion based on the features of the audio input, which were different from those collected by the SME's. The variance is likely the result of nuances in human speech that are hard to define categorically and are highly subjective to the individual. Despite this variation, the model predicted its accuracy to the speech at 89.54%, indicating a strong performance to the training dataset based on the emotional parameters defined. However, when conducted on the test dataset, accuracy in emotional recognition decreased to 85.26%. This could be a result of the male voice in the test dataset where the training was conducted with a female voice. Further explanation could be the relative tone of the pilot during the emergency, duration of the communications, and background/environmental noise during each communication. Future research would consider how to further optimise the EVR model to perform more precise and accurate analyses using a larger dataset of emotional parameters and diversity among the actors for the data collection phase. This research has applications in accident investigation roles, predominantly the identification of stressors and other hidden signs in verbal communications that can help investigators understand why the pilots made the decisions they did. There is also an application as a real-time support tool between ATCOs and pilots during high-stress and high-workload scenarios. Finally, in flight simulators, emotion recognition can be used to track how trainee pilots respond emotionally to stress, providing valuable insights into their emotional resilience and decision-making under pressure. This feedback can be used to adjust training programs, helping to improve how pilots manage their emotions and perform under critical conditions.

Conclusion

This study utilised the TESS dataset to achieve a successful EVR on the transcript between the pilot and ATCO during Flight 1549, able to capture several different emotions conveyed within a singular speech input. While the accuracy of the detection is widely disproportionate to the human-evaluated EVR, the model was able to take a highly challenging transcript and self-disseminate the emotion conveyed by the pilot. Moving forward, the study will aim to expand the TESS dataset to include a wider diversity, which will aid in the detection and classification of the AI to the emotion as represented by the human. This technology can support accident investigation during CVR transcript analyses, when identifying the pilots' cognitive and mental state before their decision-making; letting the investigator into the mind of the flight crew. Further applications can support real-time assistance during abnormal or emergencies, which could include ATCO alerting, voice-to-text conversion, and speech personalisation for high-workload environments. This area of study can also open new optimisation for future Single Pilot Operations (SiPO), providing an emotional support network between the pilot and ground network, to promote more comprehensive decision-making during emergencies, when time limitations are present.

References

Abdel-Hamid, O., Mohamed, A. R., Jiang, H., Deng, L., Penn, G., & Yu, D. (2014). Convolutional neural networks for speech recognition. *IEEE Transactions on Audio, Speech and Language Processing*, *22*(10), 1533–1545. https://doi.org/10.1109/TASLP.2014.2339736

Alharbi, S., Yahya, I., Alsarhan, A., Sedeeq, E., Kareem, A., & Hussein, M. (2021). Automatic speech recognition: Systematic literature review. *IEEE Access, 9,* 131858–131876. https://doi.org/10.1109/ACCESS.2021.3112535

Barhoumi, C., BenAyed, Y. (2025). Real-time speech emotion recognition using deep learning and data augmentation. *Artificial Intelligence Review*, *58*(2). https://doi.org/10.1007/s10462-024-11065-x

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences, 114*(38), E7900–E7909. https://doi.org/10.1073/pnas.1702247114

Ittichaichareon, C., Suksri, S., & Yingthawornsuk, T. (2012). Speech Recognition using MFCC. *International Conference on Computer Graphics, Simulation and Modeling (ICGSM'2012)*. https://www.academia.edu/download/85360099/9_20712576.pdf

Luo, F., Guo, W., Yu, Y., & Chen, G. (2017). A multi-label classification algorithm based on kernel extreme learning machine. *Neurocomputing*, *260*, 313–320. https://doi.org/10.1016/j.neucom.2017.04.052

Sherstinsky, A. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. *Physica D: Nonlinear Phenomena*, *404*. https://doi.org/10.1016/j.physd.2019.132306

Zhao, J., Huang, F., Lv, J., Duan, Y., Qin, Z., Li, G., & Tian, G. (2020). Do RNN and LSTM have Long Memory? In H. D. III & A. Singh (Eds.), *Proceedings of the 37th International Conference on Machine Learning* (Vol. 119, pp. 11365–11375). PMLR. https://proceedings.mlr.press/v119/zhao20c.html

Exploring the Perception, Challenges and Benefits of Cabin Crew Peer Support Programmes

Jordan M Hazrati & Rebecca L Grant

Coventry University, UK

SUMMARY

Twenty cabin crew with current operational experience across a variety of airlines and ranks participated in semi-structured interviews, to determine their views on the perception, challenges and benefits of peer support programmes. Perceptions of peer support were generally positive and involved talking to colleagues about a difficulty that was being experienced to feel 'heard', 'listened to' and supported by those who understood the lifestyle. The potential benefits of peer support were clear, however, concerns surrounding the trust and confidentiality of colleagues, as well as barriers to success such as lack of trust of management, financial instability of the industry and organisational culture were highlighted. Findings include recommendations to research the issues quantitively with a more diverse group of cabin crew and to industry to implement bespoke, externally–managed programmes, supported by management.

KEYWORDS

Peer support programme, cabin crew, mental health

Introduction

Cabin crew are important, frontline employees within the aviation industry (Tsaur, 2020). Whilst they also provide the onboard service to passengers, their primary purpose is to ensure the safety of the cabin, passengers and crew who travel within it. Cabin crew have not always been offered satisfactory supervision and social support, with Hajiyousefi et al suggesting that only half of flight attendants surveyed within a study conducted in Norway were satisfied with supervision and support (Hajiyousefi, 2016). Social support is thought to be able to improve safety behaviours of cabin crew (Liu, 2022), with support recognition from different sources such as colleagues stabilising the three-way interaction between proactive personality, safety climate and social support (Liu, 2022).

Supporting workers with mental health not only is an 'ethical, legal, and business benefit' but also improves 'productivity, morale, workplace behaviours, communication and decision – making' (Santilhano, 2019, p.67). Managing stress and supporting mental health is paramount to ensuring flight safety with Dismukes et al stating that stress can adversely affect the cognition and skilled performance of pilots as well as experts in other domains (Dismukes, 2015, p.1), exploring the categories of attention, working memory, decision making, team performance, and communication (Dismukes, 2015).

Aviation personnel face significant challenge and stressors as a result of industry and occupational changes; adaptive coping strategies are recognised as a component of therapeutic interventions for work-related stress (Cahill, 2021a) one example of which is the introduction and promotion of social support such as peer support programmes (Cahill, 2021a). Furthermore, stress is described as an 'insidious threat' to aviation safety, due to the 'impairments in alertness and performance it creates' (Hajiyousefi, 2016, p.32). The lifestyle for cabin crew can vary, dependent on the type of flying undertaken, base location, and the onboard rank. All cabin crew are likely to experience working unsociable working hours, periods of isolation, and fatigue/tiredness. In addition, cabin crew 'hassles' within their work environment include job, peer interactions, passenger service, and personal hassles (Tsaur, 2020, p.1). McNeely et al suggest that cabin crew experience higher prevalence of 'fatigue, depression, anxiety, and sleep disorders, as well as reproductive and all cancers' (McNeely, 2018, p.4), which increase the likelihood of mental health conditions compared to the general population (McNeely, 2018, p.9).

The aim of peer support is to offer those that need, a confidential service where they can speak openly and honestly about their problems, receiving signposting where required. The service within the aviation industry uses trained volunteers from within the airline, supported by aviation psychologists. Peer support programmes are one tool available to support people with their mental health and wellbeing, providing individual, operational safety (Yelgin, 2021) and organisational benefits.

Understanding the regulation, and current climate is important to determine the purpose and potential benefits of Peer Support Programmes. One response to the Germanwings flight 9525 incident was the introduction of regulation (EU) 2018/1042, now (EU) 2020/745, mandating the implementation of a support programme for pilots. In order to comply with regulation, airlines must provide a peer support programme to pilots (EPPSI 2019), the aim of which is to support pilots with their mental health and wellbeing (EPPSI, 2019, CAA, 2018).

EASA suggest that positive feedback was provided on the proposal regarding support programmes from commentators upon introduction (EASA, 2016, p.25), with benefits including an improvement in under-reporting, increased awareness of mental health and wellbeing, and a more open culture regarding the use/misuse of psychoactive substances or other psychological/psychiatric problems (EASA, 2016). More recently, the attention and focus within the industry has turned to whether a similar initiative for supporting the mental health and wellbeing of cabin crew should be regulated and implemented, given similar issues may be prevalent in the cabin crew population (whilst noting there are professional differences), and whether there would be improvement to operational cabin safety (Tsaur, 2020).

Cahill et al, discuss whilst peer support programmes are in place for the pilot community, they remain an unused service likely due to the stigma surrounding loss of license/medical, but also due to existing supports being highlighted as unfit for purpose (Cahill, 2021a, 2021b). Moreover, the research highlighted how there was a desire for peer support programmes for all members of the aviation workforce, and not just pilots with 75.9% of cabin crew questioned responding that they would use a peer support programme if there was one in place. However, the underuse of existing services for pilots, suggests that there is a disconnect between the desires of the workforce, and the reality of the peer support programme infrastructure.

The challenges faced by cabin crew, highlights the importance of the topic of mental health and wellbeing with this occupational group. There is currently limited research as to the perception of peer support in general and the impact of peer support programmes for cabin crew. Therefore, the aim of this work was to explore the perception, challenges, and potential benefits of peer support programmes amongst cabin crew.

Method

Design

Semi-structured interviews were chosen to ensure that whilst interviews focused on the necessary themes to gain relevant information, sufficient flexibility was offered to answer openly, whilst also providing perspectives that may not have been previously identified (Coolican, 2014). An interview schedule was developed drawing on relevant literature.

Sample and Participants

A sample of current, operational main crew and onboard managers, from different airlines within the UK was sought. A variety of cabin crew from short-haul, leisure, private and long-haul operations as well as varied service lengths in the industry were consulted to provide a representative sample.

Ethics approval

Ethical approval was given by the university ethics committee and British Psychological Society ethical procedures were followed (British Psychological Society, 2018).

Procedure

Interviews were conducted over Microsoft Teams. Interviews commenced by reviewing the demographics of participants, and discussing their experiences within the industry, helping to build rapport and trust. Following this, the perceptions, challenges and benefits of peer support were explored. Interview discussion points included suitability for the role, skills and training, confidentiality, supporting the peer supporter, implication to safety, and overall wellbeing. Appropriate semi-structure interview techniques were utilised in line with Kvale and Brinkmann's guidance (Kvale, 2018). Interviews lasted between 27 and 63 minutes.

Analysis

Data were thematically analysed via NVivo, using Braun and Clarke's reflexive thematic analysis. This identified, coded, and analysed emerging patterns and trends within the interview responses, in a detail rich, yet unrestricted way, whilst acknowledging the centrality of researcher subjectivity and reflexivity (Braun, 2019, 2020). This also included an analysis of literature and regulatory documents previously explored to guide initial coding themes. Data gathered were then transferred to quote expansion tables.

To ensure quality of the research Lincoln and Guba (Lincoln, 1985) trustworthiness checks were conducted within the categories of credibility, transferability, dependability, and confirmability. This process took place prior to, during and post – data collection, including incorporating the use of an SME at the initial interview schedule creation, thematic evaluation and report production phases.

Results and Discussion

Sample

Participants held a wide range of experience across differing airline operation types and airlines, eight with short and long-haul experience, nine long-haul only, one short-haul only, and two with private aviation experience. The overall average length of service was 14.05 (SD = 9.88) years, with eight On Board Manager crew averaging a 19-year career (SD = 10.50), and 12 Main Crew averaging a career of 10.75 years (SD = 8.28).

Research Themes

The results and discussion are split into three separate themes (with sub-themes). The three main themes were perception, challenges and benefits, as shown in Table 1, and are illustrated via short, abbreviated quotes focussing on key words in line with Merriam (Merriam, 2009).

Table 1: Research Themes

Theme	Category
Perception	Perception of peer support and/or peer support programmes. Perception of suitability for the role of peer supporter of cabin crew. Perception of required skills and training. Trust of cabin crew.
Challenges	Challenges of a peer support programme. Confidentiality. Support the peer supporter. Management support.
Benefits	Purpose (benefits) of peer support. Safety. Wellbeing.

Perception

This section explores the theme of perception, focusing on the support programme, the role of peer supporter, as well as skills, training and trust of cabin crew.

Perception of Peer Support and/or Peer Support Programmes

The perceptions of peer support were mixed and based upon a variety of factors. One participant stated that when they thought of peer support they thought *'of the channels that the company has set up that help with mental health'* (P3) rather than a specific peer support programme, indicating that they were unsure of what a peer support programme was. Others said that peer support was *'talking to your peers, your colleagues and being helped by them'* (P1) and *'speaking to someone'* (P5) summarising that *'peers are friends'*(P11), or *'somebody in the same position as you,'* (P14).

There were some participants that had *'never'* (P17) heard of peer support, however thought *'it would be an amazing thing'* (P17), and that *'airlines should have something in place'* (P11). The lack of knowledge, but desire for support suggests that whilst some airlines do not have a dedicated peer support programme, those that do, likely do not advertise or signpost their programmes regularly to cabin crew supporting the views of Cahill et al (Cahill, 2021a). However, it was also noticed that whilst mental health has become a more openly spoken about and accepted concept, *'it's still got a stigma attached to it',* (P4) due to the concerns surrounding loss of medical or loss of career, mirroring the views expressed by Cahill et al (Cahill, 2021b) and the pilot community. Peer support was viewed by participants as suitable for *'everyone'*, as 'everyone has issues with their mental health' (P3), and that 'mental health has a significant impact on cabin crew' (P3) building on the findings from Tsaur et al in relation to the hassles experienced by operational crew (Tsaur, 2020).

Perception of Suitability for the Role of Peer Supporter of Cabin Crew

Participants thought that *'only crew really understand crew'* (P6), and it was felt that *'crew members will pull together'* (P13) and support each other. Some job factors are very unique to the cabin crew role, making it difficult for a peer supporter to empathise if they haven't experienced it for themselves, for example loneliness and feeling of missing out on events at home, representing views regarding job hassles expressed by Tsaur et al (Tsaur, 2020) and the unique challenges that the aviation industry presents for cabin crew.

A participant explained that cabin crew were suitable because whilst they all came from *'different backgrounds'*, they were *'also similar'*, and that *'by the end of your eight-hour sector or whatever it is you're like, a little family'* (P2). This for them meant they already had the ability to trust and confide in cabin crew, so would feel comfortable approaching them for peer support. However, one participant explained that they held concerns fearing *'they might not feel comfortable talking to them because they know them'* (P6), contrasting Tsaur et al (Tsaur, 2020) and McNeely et al (McNeely, 2018) .

Variety in terms of age, experience, and demographic within peers was deemed important to promote use, as would *'a 58-year-old going through something [be] happy talking to an 18-year-old?'* (P18). The experience of some cabin crew who have *'been crew for like 20-30 years'* (P9) was seen as advantageous, as they would be able to relate to a range of experiences, empathising with the unique hassles of the role discussed by Tsaur et al (Tsaur, 2020).

Perception of Required Skills and Training

There was a general agreement that cabin crew naturally had a lot of the soft skills required to be peer supporters, for example the ability to be *'empathetic and sympathetic'* (P6) and are *'caring'* enough to *'speak to people and understand and just be open and honest'* (P20). This would help to build rapport and trust between a peer supporter with core principles such as 'empathetic listening key to success of a peer support programme' (Santilhano, 2019, p.73). These soft – skills are required to fulfil the competencies of operating as cabin crew and are therefore likely to be naturally found within this community.

Training in how to signpost to, and support from *'professionals'* (P13) was also perceived as an important skill to learn. Cabin crew who had previous experience were perceived as valuable with one participant expressing that they would feel comforted if their peer supporter had *'been an NHS nurse, or psychiatric carer'* (P16), had undertaken *'mental health first aid'* (P17) or even had *'personal experience of maybe dealing with situations'* (P8). This is useful experience as the peer supporter will likely have increased knowledge and understanding of mental health, and appropriate forms of signposting, as well as confidence in their ability to support cabin crew, leading to a higher chance of a successful peer support programme interaction as discussed by Santilhano et al (2019).

Trust of Cabin Crew

The perception of whether cabin crew would make appropriate peer supporters, and thus the success of a peer support programme was associated with whether participants felt as though they could trust their colleagues; *'could you ever fully trust them?'* (P1) one participant mentioned, suggesting that some cabin crew were seen as *'snakes in the grass'* (P1), and would be using their position as a peer supporter for ill-intentioned means, for example information *'getting back to the manager'* (P2), supporting Cahill et al's research highlighting the necessity, but not guaranteed factor of trust to the success of a peer support programme (Cahill, 2021a). A participant explained that they *'wouldn't want to go and then be grassed up'* (P2), with concerns shared surrounding whether the peers would *'keep it to themselves'* (P8). The fear of repercussion from management was mirrored in several testimonies with one participant disclosing that *'I know people who have been contacted by ground management and they've said we know you've done this'* (P1), which has dissuaded cabin crew from contacting support in the future. This view was not supported by all participants, with most supporting the desire for an external, confidentially run service, but desiring the relatability of speaking to fellow colleagues.

Challenges

Interviews explored the challenges of implementing peer support programmes, based upon perception and existing mental health concerns.

Challenges of a Peer Support Programme

Organisational challenges and barriers were cited by participants in the implementation of a peer support programme or improvement of an existing one. The first challenge related to financial impact with phrases such as *'airlines are very reluctant at the moment to spend money'* (P1), *'cash is king',* (P2) and *'cost versus value benefit'* (P12) mentioned. Convincing management committees and senior leadership that implementing a programme that will potentially incur additional cost, is perceived to be a challenge, however potentially of benefit to the safety climate of an organisation as discussed by Liu et al (Liu, 2022).

Another challenge is the *'resourcing'* (P2) required to run a peer support service, with one participant asking, *'why would they take you out of the operation?'* (P2), to train and volunteer as a peer supporter, particularly poignant given the reduction of personnel in recent years. Accessibility was another perceived challenge by participants, with a peer support programme needing to be *'easily accessible'* (P1) to have any impact. In addition, *'time putting it together'* (P11), as well as the concern surrounding *'how long is it gonna take to roll out?'* (P18) was cited. Participants worried *'whether the airline would promote it?'* (P13) with commitment required as discussed by Santilhano et al from management regarding time and resource for mental health and wellbeing initiatives in order to improve trust in the programme (Santilhano, 2019).

Confidentiality

One of the most important functions of a proposed or existing peer support programme for participants was confidentiality. A lack of confidentiality was cited as a concern or barrier to the success of a peer support programme. It was described that *'trust is very, very hard to gain but very easy to lose'* (P19), and there was a *'fear of someone telling'* (P9) described by another participant. Confidentiality therefore is viewed as important because *'you want people to feel comfortable and safe'* (P5), especially when talking about subjects such as mental health that as stated by Cahill et al, have perceived stigma attached to them (Cahill, 2021b).

There was discussion by all participants about when confidentiality should be broken, with an unanimously agreed upon view expressed that whilst any programme has to be *'100% confidential there has to be a point in which the advisor decides is this person fit to be on an aircraft as operating crew?'* (P16) including if it was deemed that the colleague using peer support might *'harm [themselves] or harm others'* (P6). Therefore, it can be determined that if a peer supporter felt that the safety of the individual or the operation was at risk, there should be an escalation process to support both the operation and the individual, breaking confidentiality.
A participant also discussed the potential for removal of confidentiality should an emerging trend be identified amongst the peer support programme. The participant stated that *'if it could be a potential safety issue then that's when it needs to then be reported to the company'*(P20). However, this would have to be handled sensitively, as discussed by Santilhano et al (2019), confidentiality is critical to the integrity of a peer support programme and could be extremely difficult to operationalise.

Support the Peer Supporter

It is *'important [to have] a safe place for the councillor or the peer supporter, the listener to be* *able to offload'* (P14). This was explained as useful because what the peer supporter may experience could be upsetting and provide a challenge to the programme with attracting and retaining peer supporters, and therefore a secondary effect of volunteering to be a peer supporter could potentially be a decline in personal mental health, should it not be supported appropriately, further heightening the stress caused by job hassles mentioned by Tsaur et al (Tsaur, 2020) and mental health conditions highlighted by McNeely et al, experienced by cabin crew (McNeely, 2018).

Management Support

The success of a peer support programme was attributed by many participants to the support from management. One participant commented that a barrier to the success of a programme was *'management and [the] organisation'* (P2), with it being felt that a peer support programme could just be used as a way of *'ticking boxes'* to make it *'look like [they're] doing something'* (P2). The lack of trust in the relationship between cabin crew and management was a reciprocated dynamic with one participant stating that *'crew don't trust management, management don't or don't seem to trust crew'* (P13). To improve that, one participant suggested that *'as the CEO of a company, you've really gotta care about your company'* (P11), and by implementing a peer support programme, this would be a step closer to improving that dynamic between the workforce and senior management, supporting the view researched by Santilhano et al (2019) that by supporting mental health, there would be improvement to productivity, morale and workplace behaviours.

One way recommended to build the trust of the programme involved management communication and endorsement with one participant stating that *'for this to work, the management would not only need to market it, but they need to be the ambassador for it'* (P20). However, many cabin crew would prefer a programme to be kept separate from management to encourage trust, and therefore it is concluded that whilst positive endorsement, and encouragement of a peer support programme is essential to success, the service should remain external to the company and from management, ensuring confidentiality supporting the findings of Cahil et al (2021a) and Santilhano et al (2019).

Benefits

Research within this paper is presented from both an organisational and individual perspective.

Benefits (purpose) of Peer Support

Participants had a range of ideas regarding the purposes and subsequent benefits of a peer support programme. One participant explained that from their experience with an existing programme it was quite important to have somebody to talk to who might not be your closest family or friends, as *'you might not want to talk to them'* (P1) nor might you want to speak to a professional about something. Overall, participants felt as though the purpose of peer support programmes should be to talk about *'anything that's worrying you'* (P8), and that it should be a place where *'crew can feel supported'* (P2) with their mental health and wellbeing.

Participants commented that the programme could be used for a range of issues including *'work-related'* (P1), *'home-related'* (P3), *'financial'* (P3) and for serious issues at work such as *'death on board'* (P6), *'mental-health concerns'* (P13) and *'demands of the lifestyle'* (P15).

Participants expressed that it should not always need to be used for extreme events and should be used to *'support crew in their day to day lives and well as when it gets a little bit more serious'* (P3). This was seen as normalising the conversation surrounding mental health, removing the stigma discussed in research from Cahil et al. This develops the organisational culture of the airline, by encouraging a psychologically safe environment, in which discussion surrounding mental health and performance variability was supported, supporting the views of Liu et al (Liu, 2022) in improving the safety climate of an organisation. Overall, it was felt that the purpose of a peer support programme should be to *'support'* (P10) within a non-judgmental, *'no-blame culture'* (P8).

Safety

The most important part of a cabin crew member's role is to ensure the onboard safety of passengers and crew travelling within the cabin as discussed by Tsaur et al (Tsaur, 2020) therefore it is important to ascertain whether there would be improvement to operational safety. It was thought that there would be an *'impact on safety positively'* (P18). Another participant suggested that *'the highest performing teams will be the safest teams'* (P12), with words such as *'safer, nicer environment'* (P5) and *'open kind of culture'* (P1) used to explain how the improvement to safety would manifest amongst cabin crew. It was felt that peer supporters could not only preserve the career of the individual, but the safety and security of an airline by supporting cabin crew to consider reporting unfit for duty, supporting the view of Dismukes et al (Dismukes, 2015).

Where wellbeing is not supported, *'people create work arounds, and sometimes you don't always do what the manual says to the word'* (P2). Some participants also felt that there would be an improvement to safety culture. It was expressed that cabin crew were *'more likely to own up to potential things that have happened'* (P1) if appropriate support was in place, with another view that they felt the *'psychological contract'* (P2) between the cabin crew member and the airline would be improved if wellbeing was prioritised, as cabin crew would want to focus on the integral safety part of their role, in line with Liu et al (Liu, 2022) and the Aerospace Medical Association (Aerospace Medical Association, 2021).

Wellbeing

'General morale' (P16) was mentioned by participants as a benefit, as well as *'less rates of sickness'* (P20) and a sense of relief, and *'possibly [being] happier going into work'* (P7). A participant described how they thought that *'when we don't feel listened to or supported, it can make you feel over - emotional, and I think mistakes can get made as well'* (P14). Another supported this stating that *'if your mental health is not there, if you are feeling with that in any way shape or form, it's exceptionally hard to open yourself up to brand new people'* (P12). One participant felt that their airline *'could be a champion for this, and we can make this a better place to work where people want to come and work'* (P7), which would improve recruitment strategies, as well as the individual wellbeing of cabin crew when considering the unique stressors associated with the role discussed by Tsaur et al (2020) and McNeely et al (2018).

Ultimately, it was thought that a *'[peer support programme] will make people happier'* (P5), *'problems aren't likely to get exacerbated'* (P6) and a peer support programme will *'give us support, so coping strategies and how to get through it'* (P6). This is ultimately because it is perceived that *'mental health in the crew community needs to be addressed'* (P10), and *'the sooner people get on board with something, then it will be better for all of us'* (P7).

Conclusion

The key research findings are as follows:

(1) Peer support programmes are considered to be a positive initiative to support the mental health and wellbeing of cabin crew.

(2) The implementation and management of a peer support programme should be external to the company, to promote confidentiality and ensure trust.

(3) The value of senior leadership is paramount to programme success, specifically via meaningful promotion, support and encouragement of service use.

(4) Cabin crew have many of the relevant transferable non-technical skills that would make them strong candidates for the role of peer, however diversity amongst peers and training and support for supporters is paramount.

Recommendations

(1) Following the identification of key perceptions on the issue, these themes should now be investigated quantitatively on a wider sample of cabin crew.

(2) Airlines should consider implementing a peer support programme for cabin crew, involving the end user during design phases to maximise suitability.

(3) If a programme exists, it is recommended that further research is done within the airline to determine the current usage, and whether the programme remains fit- for-purpose.

References

Aerospace Medical Association. Covid-19, Aviation Personnel and Mental Health Support. 2021. http://www.asma.org/asma/media/AsMA/Travel-Publications/Pilot%20Mental%20Health/Coping-with-Covid19-2-19-2021.pdf [Accessed 10th January 2022].

British Psychological Society. 2018. Code of Ethics and Conduct. available from https://www.bps.org.uk/news-and-policy/bps-code-ethics-and-conduct [Accessed 20 January 2022].

Braun V, Clarke V. One Size Fits All? What Counts as Quality Practice in (Reflexive) Thematic Analysis. Qualitative Research in Psychology. 2020; 18(3).328-352.

Braun V, Clarke V. Reflecting on Reflexive Thematic Analysis. Qualitative Research In Sport, Exercise and Health. 2019;11(4).589-597.

Cahill J, Cullen P, Answer S, Gaynor K. 2021a. White Paper: Impact of COVID 19 Pandemic on Aviation Workers and the Aviation System. COVID Survey White Paper, Centre for Innovative Human System (CIHS). 2021. Trinity College Dublin: Ireland.

Cahill J, Cullen P, Answer S, Wilson S and Gaynor K. Pilot Work Related Stress (WRS), Effects on Wellbeing and Mental Health, and Coping Methods. The International Journal of Aerospace Psychology. 2021b; 31(2):87-109.

Civil Aviation Authority. Pilot Support Programme. Guidance for Commercial Air Transport (CAT) Operators. 2018.

Coolican H. Research Methods and Statistics in Psychology. 2014. 6th ed. Psychology Press.

Dismukes R K, Goldsmith T E, Kochan A. Effects of Acute Stress on Aircrew Performance: Literature Review and Analysis of Operational Aspects. 2015. NASA: California.

EASA. Aircrew medical fitness: Implementation of the recommendations made by the EASA- led Germanwings Task Force on the accident of the Germanwings Flight 9525. (Changes to Regulation (EU) No 965/2012). 2016.

EPPSI. Pilot Peer Support Programmes: The EPPSI Guide. 2019. EPPSI-Guide-v8.1.pdf.

Ford J, Henderson R, O'Hare D, 2014. The effects of Crew Resource Management (CRM) training on flight attendants' safety attitudes. Journal of Safety Research. 2014;48:49-56.

Hajiyousefi H, Asadi H, Jafari A. Work Stress amongst Flight Attendants: The Perspective of a Standard Sports Examination Designing as a Prerequisite to Flight License. International Journal of Sciences: Basic and Applied Research. 2016; 11(1). 31-40.

Kim H, Yu M, Hyun SS. Strategies to improve work attitude and mental health of problem employees: Focusing on Airline Cabin Crew. International Journal of Environmental Research and Public Health. 2022;19(2):768.

Kvale S, and Brinkmann S. Doing Interviews. 2018. (2nd edition). London: Sage.

Lincoln Y, Guba E. Naturalistic inquiry. 1985. Newbury Park, California. London: Sage.

Liu B, Quan X, Zin X, Cui X, Ji M, You X. How can proactive personality affect cabin attendants' safety behaviours? The moderating roles of social support and safety climate. International Journal of Occupational Safety and Ergonomics. 2022; 29(1). 243-253.
McNeely E, Mordukhovich I, Tideman S, Gale S, Coull B. Estimating the health consequences of flight attendant work: comparing flight attendant health to the general population in a cross-sectional study. BMC Public Health. 2018; 18(346).
Merriam S. Qualitative Research: A Guide to Design and Implementation. 2009. Wiley: Somerset.
Mulder S, de Rooy D. Pilot mental health, negative life events and improving safety with peer support and a just culture. Aerospace Medicine and Human Performance. 2018;89(1). 41-51.
Santilhano W, Bor R, Hewitt, L M M. The role of peer support and its contribution as an effective response to addressing the emotional well-being of pilots. Aviation Psychology and Applied Human Factors. 2019; 9(2). 67-76.
Tsaur S-H, Hsu F-S, Kung L-H. Hassles of cabin crew: An exploratory study. Journal of Air Transport Management. 2020;85:101812.
Yelgin Ç, Ergün N. The effects of job demands and job resources on the safety behavior of cabin crew members: A qualitative study. International Journal of Occupational Safety and Ergonomics. 2021;28(3):1511–21.

HAIQU - A Human Factors Requirements App for Human-AI Teams in Aviation

Venditti, R.[1], Pozzi, S.,[1] Frau, G.,[1] Salam, R.,[1] Imbert, J-P.,[2] Duchevet, A.[2] and Kirwan, B[3]

[1]Deep Blue, [2]ENAC, [3]EUROCONTROL

SUMMARY

Contemporary Human Factors requirements sets, such as those embodied in EASA CS29.1302 for cockpit design, and the SESAR Human Performance Assessment Process (HPAP) requirements for air traffic management systems, are unlikely to be sufficient to account for Operational Explainability (OpXAI), shared situation awareness, and other elements associated with proposed Human-AI Teaming systems. Whilst the European Union Aviation Safety Agency (EASA) has provided new guidance on a number of these areas, their focus is largely on safety, with less attention on other areas such as Roles and Responsibilities, Competencies and Training, and Organisational Readiness - all concerned with systems integration.

HAIQU (Human-AI Teaming Questionnaire) is a freely available Web App developed in the Horizon Europe HAIKU project. The app aims to make standards and regulations accessible and user-friendly for design teams looking to integrate AI capabilities into safety critical applications. HAIQU contains 180 requirements in eight Human Factors areas (Human-Centred Design, Roles & Responsibilities, Sense-Making, Communication, Teamworking, Error and Failure Management, Competencies and Training, and Organisational Readiness). The App is sensitive to different design maturity levels and AI autonomy levels, consistent with EASA's classification of Human-AI Teaming arrangements.

This paper firstly situates the requirements set in terms of current EU and EASA regulations, as well as contemporary Human Factors guidance sources for cockpit and air traffic management systems. It then focuses on showing how HAIQU works and can serve as an aid to Product or Design Teams managing the integration of Human Factors into future AI-based systems, using as illustration a use case in cockpit Human-AI Teaming.

KEYWORDS

Aviation, AI, Human-AI Teaming, Human Factors Requirements, Design

Background

The next decade is likely to see the introduction of AI-based Intelligent Assistants (IAs) in operational aviation contexts, whether to augment pilot, controller, and airport operatives' capabilities, or to support new concepts such as single pilot operations and urban air mobility (city-wide drone and sky-taxi air traffic control)[1]. Whilst some argue that AI is 'just more automation' (e.g. Kaliardos, 2023), others consider that future AI could herald a more radical change in human machine collaboration. AI has the potential to drastically impact several sectors, improving different key performance areas, i.e. safety, efficiency, predictability. However, it also presents a

[1] See https://haikuproject.eu and https://safeteamproject.eu

new set of challenges such as algorithmic bias, lack of transparency, over-reliance and unclear human-machine responsibility boundaries.

The European Union's Artificial Intelligence Act (AI Act: European Parliament, 2023[2]) introduces a proportionate risk-based approach to AI regulation, which imposes a gradual scheme of requirements and obligations depending on the level of risk posed to health, safety and fundamental rights. This approach assigns regulatory requirements to four distinct risk categories: unacceptable, high, limited and minimal (see Figure 1). This hierarchical structure follows the principle that higher risks warrant stricter requirements and obligations, aiming to balance innovation with the protection of fundamental rights and values.

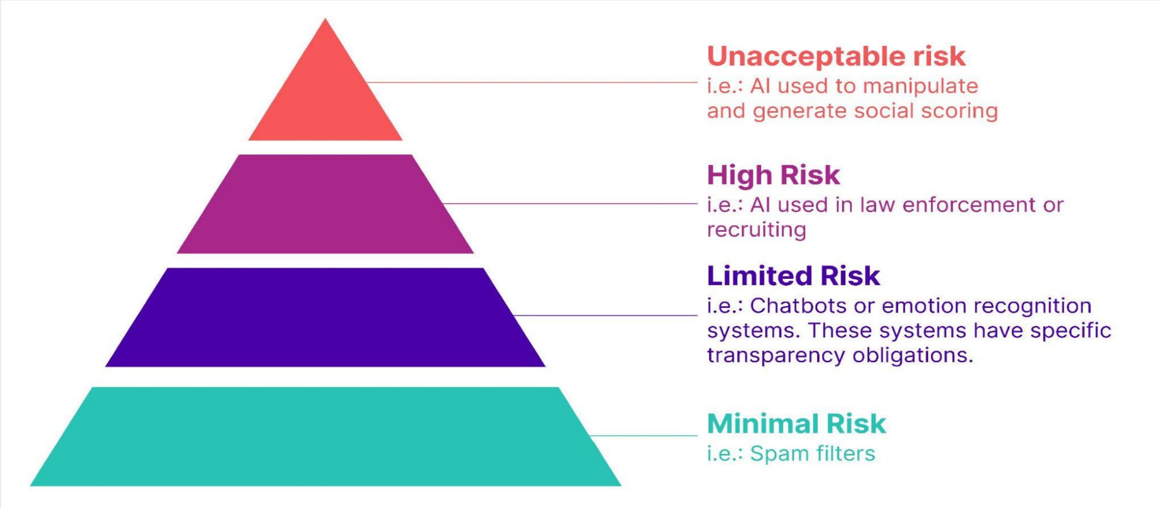


Figure 1: A Proportionate Risk-based approach to AI (adapted from European Parliament, 2023)

The Need for a Requirement Tool

The EU AI Act requirements complement other aviation sector-specific regulations, such as those established by the European Union Aviation Safety Agency (EASA, 2023; 2024). EASA has developed guidance on Human Factors for Human-AI Teaming systems, encompassing operational explainability, natural language processing interfaces, and traditional aspects like design and human-machine interaction, bringing new requirements to the scene. However, EASA's focus is safety, whereas Human Factors (HF) considers in more depth other performance elements such as roles and responsibilities, competencies, and wellbeing, that can have indirect impacts on safety.

Some of these factors are highlighted in the European Act on AI, but not crystallised into measurable requirements. HF requirements for AI systems therefore remain somewhat scattered across different sources and often lack practical implementation guidance for designers. Additionally, while regulatory entities effectively communicate principles and requirements through declarative statements ("the designer shall"), they rarely address the procedural aspects ("how to").

The HAIQU (Human-AI Questionnaire) Web App addresses this gap by translating regulatory requirements into actionable questions for design teams, helping them improve their AI systems and meet compliance standards. HAIQU has been developed in the context of the Horizon Europe funded HAIKU (Human-AI teaming Knowledge and Understanding for aviation safety) project

(https://haikuproject.eu). HAIKU aims to explore Human-AI Teaming via six aviation use cases involving AI prototypes (2 cockpit, one ATM, two airport and one urban air mobility)

2 https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)698792

The HAIQU concept

The HAIQU team has derived and synthesized Human Factors requirements by reviewing multiple regulatory sources, including EASA's regulatory requirements for Human-AI interactive systems in aviation, as well as requirements arising from the EU Act on AI, and Human Factors guidance from platforms such as SESAR[3], EASA regulation CS25/1302[4], and the Human Factors Compass[5]. These requirements, expressed as statements, were then translated into a set of questions for Product or Design Teams to evaluate their AI design against the HF requirements relevant to both its maturity level and the level of AI autonomy in the Human-AI Teaming system concept.

As an example of how HAIQU renders declarative requirements into more achievable ones contextualised in Human Factors capabilities, in the EU AI Act it is stated that AI-based systems should be human-centred (AI Act: European Parliament, Recital 6, 2023), implying the need for human involvement in the design process. However, it provides no specific guidance on how to achieve this. The related HAIQU question is "*Are licensed end-users participating in design exercises such as focus groups, scenario-based testing, prototyping and simulation (e.g. ranging from desk-top simulation to full scope simulation)?*" With such a question the Product Team can immediately tell if the requirement is satisfied, and if not, they know what needs to be done.

Similarly, in the EASA guidance there is an operational explainability requirement (EXP 18) that "*The training and instructions available for the end user should include procedures for handling possible outputs of the ODD [Operational Design Domain[6]] monitoring and output confidence monitoring*." This is translated and expanded in HAIQU as follows:

i. "*Are users trained to recognise and take corrective action on strange or erroneous AI outputs?*"

ii. "*Does advice offered come with an indication of its confidence or uncertainty?"*

iii. *"Has the end user seen examples of AI incorrect information / advice in simulation training?"*

At a simple level these questions can be answered 'yes' or 'no'. If 'yes', evidence should be available in terms of training schedules and records [(i) and (iii) above], and of the presence of confidence estimation parameters in the AI (ii). This expansion of the requirements can give the Product Team more latitude, e.g. if confidence estimates are not available (as these can be difficult to generate with some AI systems) or are deemed inadvisable, training and testing of training in simulators [(i) and (iii)] offer viable alternatives. This points to one of the aims of the HAIKU project, namely to be flexible and proportionate, and to give the Product or Design Team options. This degree of flexibility and 'scalability' is important, as the HAIQU questionnaire is by its nature one for self-reflection by the Team – there is no set of magic or secret answers, though often the 'ideal' answer is implied in the question. Nor is it intended as a regulatory compliance tool (though it may be used to increase the likelihood of regulatory compliance). Rather, it is for the team to see

3 SESAR Human Performance Guidelines https://www.sesarju.eu/sites/default/files/documents/transversal/SESAR%202020%20%20Human%20Performance%20Assessment%20Guidance.pdf

4 EASA CS25 1302 https://www.easa.europa.eu/en/search?keys=CS25%201302&f[0]=origin:EASA+Pro

5 Human Factors Compass SAFEMODE Project https://www.safemodeproject.eu/EhuridGuidedPaths.aspx

6 Operational Design Domain defines the specific operating conditions under which an aviation system (or a part of it) is designed to function safely and effectively.

where they are strong on assuring high human-AI team performance with their system, and where more work may be advisable.

The resulting 180 questions (for more on their derivation see Kirwan, 2025) effectively guide teams into the exploration of eight key areas, *Human-Centred Design, Roles & Responsibilities, Sensemaking, Communication, Teamworking, Error & Failure Management, Competencies & Training,* and *Organisational Readiness*. An overview of the HAIQU architecture is given in Figure 2 (Kirwan, op cit).

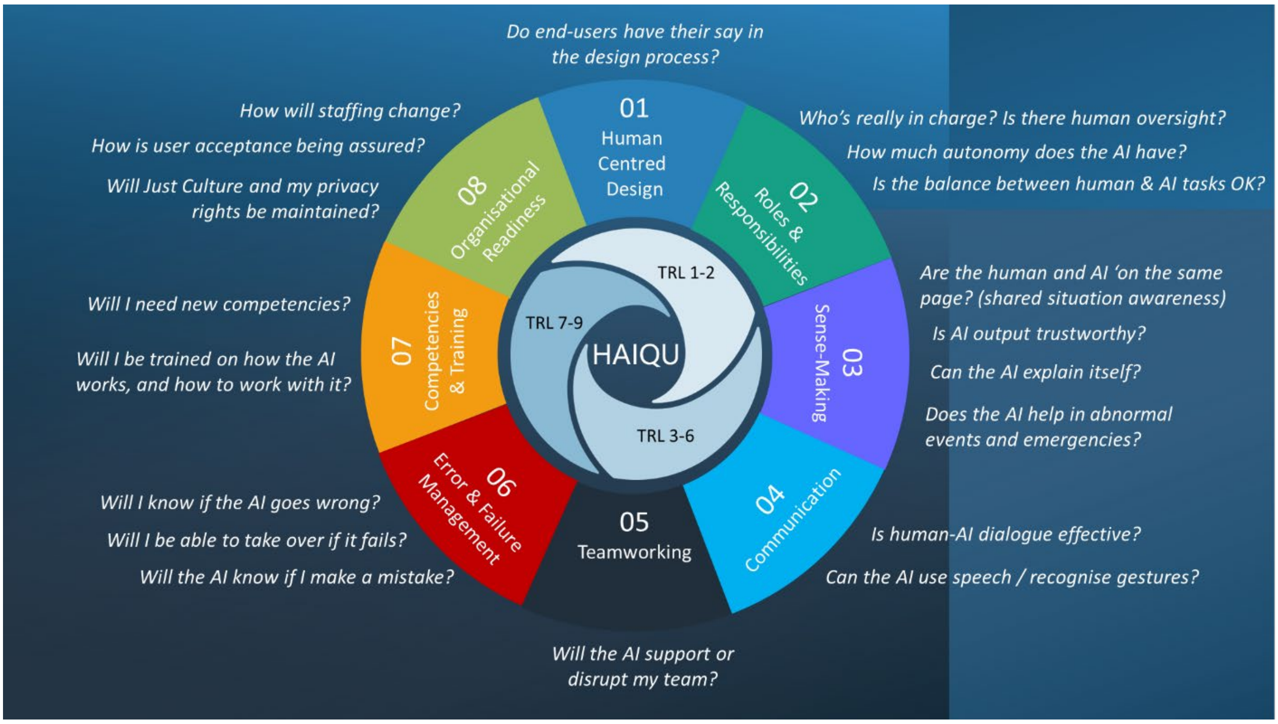


Figure 2: HAIQU Eight Overarching Human Factors Areas (from Kirwan, 2025)

How HAIQU Looks and Works

The questions are embedded in a collaborative web-platform (https://haiqu.eu/). The landing page is shown in Figure 3. Users can sign in to the platform, register their system, and begin the questionnaire immediately. HAIQU supports iterative development by enabling assessments at various stages of design maturity. While some questions may need to be revisited later in the development cycle, early identification of potential issues allows for timely adjustments before designs become too rigid to modify. By the time a system approaches deployment, teams will have comprehensively addressed all relevant requirements, establishing a solid foundation for operating their Human-AI system.

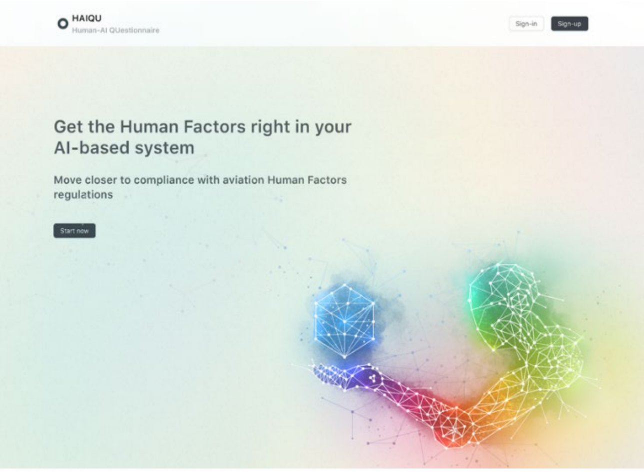


Figure 3: HAIQU Landing Page

HAIQU Features

To achieve its goal and concretely help Product and Design Teams, the HAIQU app offers several key features designed to support diverse user needs.

Progress tracking and saving: HAIQU is comprehensive, and it may not be possible to complete all relevant questions in one sitting. Some questions may be non-applicable, some will be quick to answer, and others may take more time and require deeper consideration and evidence gathering. For this reason, the tool saves the status of the questionnaire, giving the possibility for users to continue when they see fit.

Interaction with questions: Users can quickly evaluate requirements using four response options (see Figure 4): 'Yes' (requirement met), 'No' (requirement not met), 'TBD' (to be addressed later), or 'N/A' (not applicable). The interface prioritises quick navigation and fast interaction, allowing users to navigate smoothly between different assessment areas.

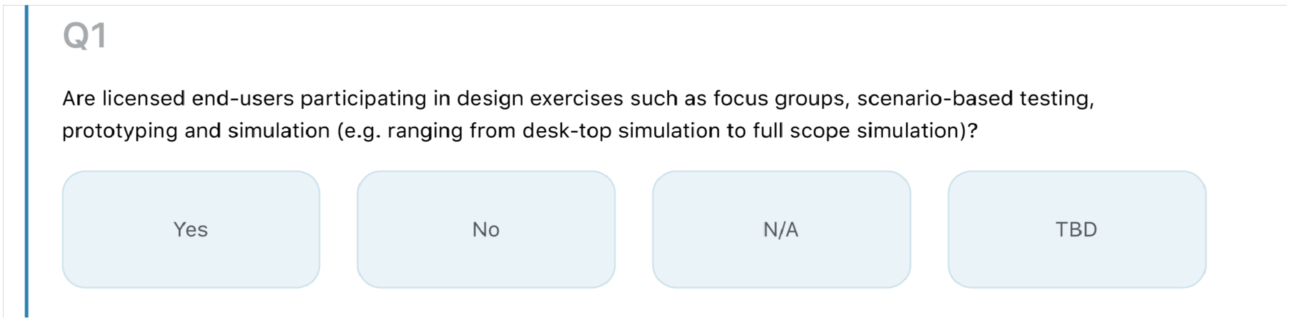


Figure 4: Example of a HAIQU Question

Action tracking: When requirements are marked as "No" or "TBD", the system prompts teams to document specific actions needed for compliance. This feature helps teams develop concrete implementation plans, with all actions visible to the entire team through a shared workspace.

Real-time monitoring: A dynamic dashboard provides real-time visibility of the evaluation progress across all eight areas or selected subsets relevant to the project. The system records all responses, including supporting evidence, justifications and identified actions. This tracking enables teams to identify areas needing attention. For instance, if a project shows strong compliance in Human Centred Design but reveals gaps in Error & Failure Management, teams can prioritise improving their error management strategies.

As an example, Figure 5 shows the HAIQU status of a 'project-in-progress' called FOCUS (Flight Operational Companion for Unexpected Situations: Duchevet et al, 2024), which is a cockpit-based research prototype Human-AI Teaming tool aimed at helping pilots in sudden in-flight emergencies.

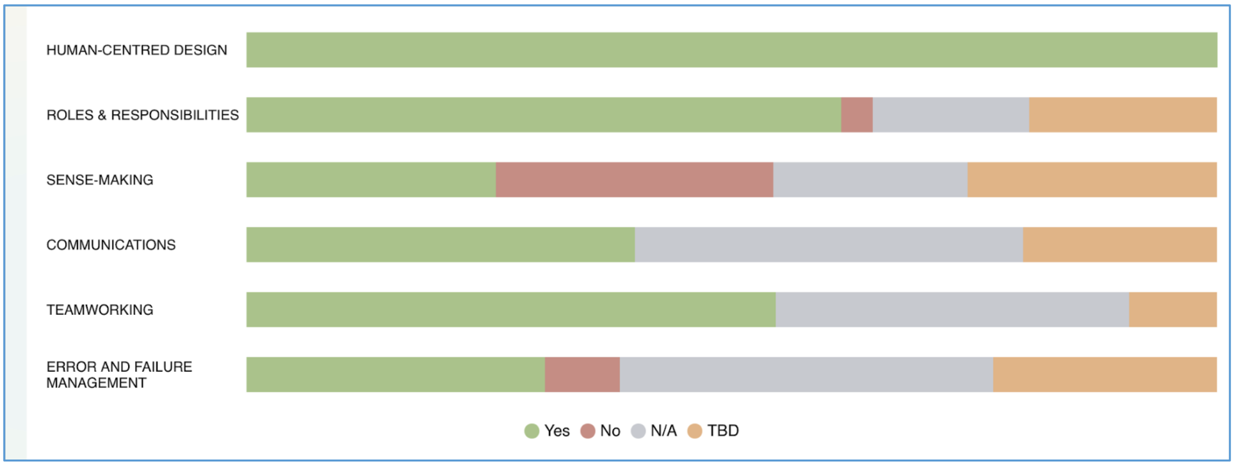


Figure 5: Example of Intermediate HAIQU Results for a Human-AI Teaming Project

This figure shows that the project is doing well in certain areas, whereas more work is needed in others. Two other parts of the Dashboard are shown in Figure 6. The left part of Figure 6 shows the status of answers to the relevant questions. For FOCUS, 120 out of 180 are deemed relevant based on the level of maturity of the concept (it is a research prototype that can be used in a real-time simulation) and the level of AI autonomy and functionality (it is EASA Category 2A and does not have AI natural language processing). The questions answered 'Yes' can point to substantiative evidence, those answered 'No' or 'N/A' (not applicable) need to have an associated justification, and those deemed TBD (To be done) need a task or work plan.

The right-hand side of Figure 6 offers a 'helicopter view' and is a spider chart showing the overall 'health' of the project when considered against the eight HAIQU Human Factors areas. In this case, due to the maturity level of the research project, *Competencies & Training* and *Organisational Readiness* are not yet assessable, so are set to zero. *Human Centred Design* is fully addressed, and *Teaming* and *Communications* are in progress, while *Sense-Making* and *Error and Failure Management* are probably the next priorities in the development process. The Dashboard can be reviewed at key stages (e.g. after a major simulation with end-users) to keep an eye on progress.

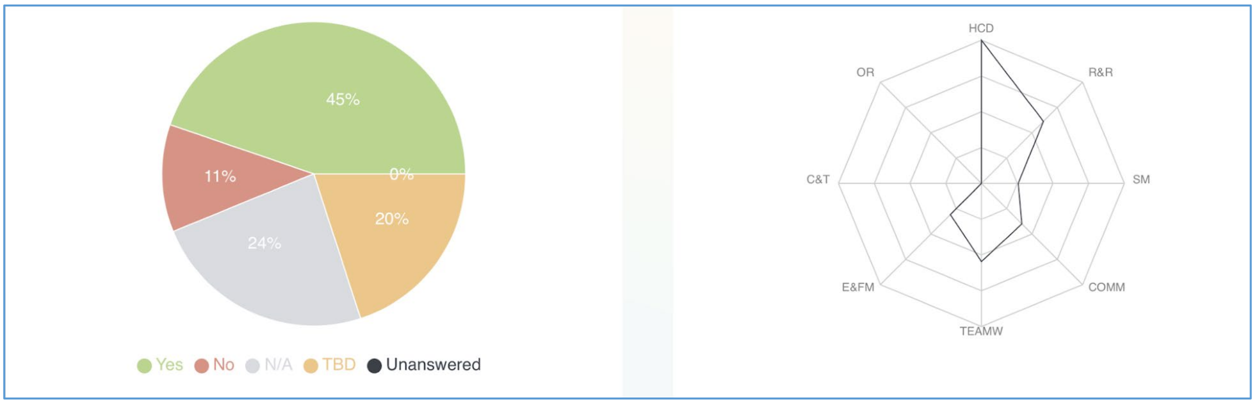


Figure 6: Key HAIQU Dashboard Elements

Figure 7 shows example 'TBD' tasks for FOCUS related to 'Sense-Making', showing the original HAIQU question and the task plan to resolve it.

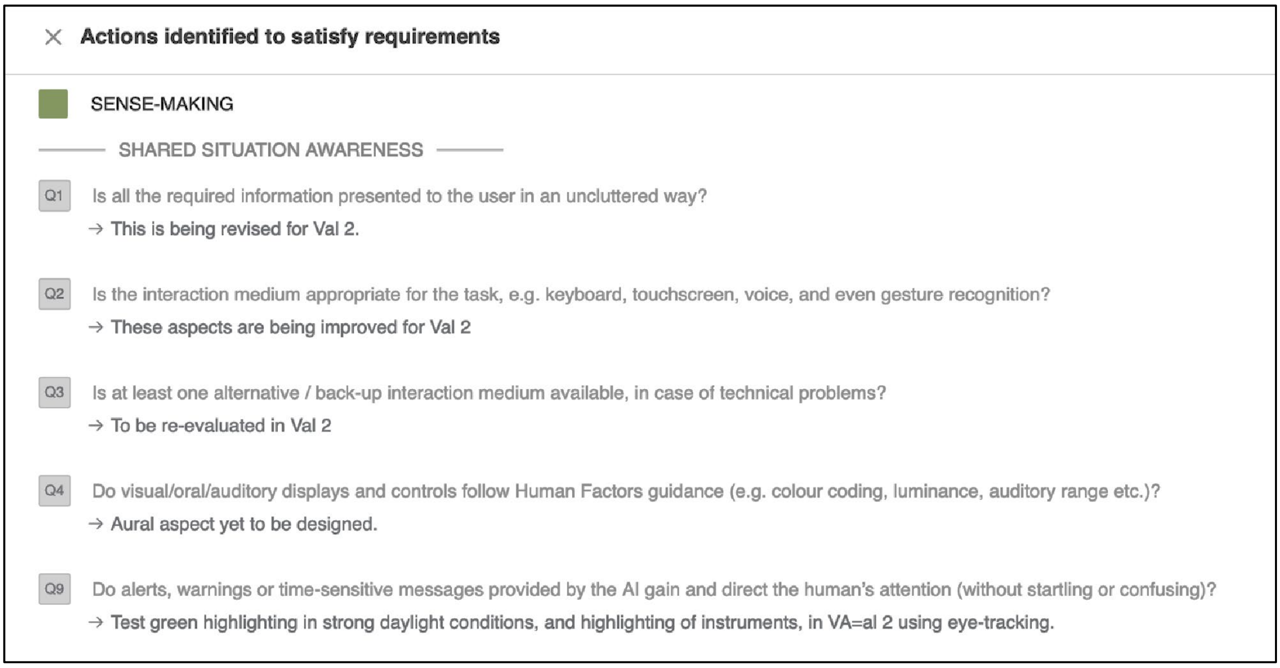


Figure 7: Example of residual tasks to be carried out to satisfy requirements.

In Figure 7, 'Val 2' refers to a second real-time simulation with pilot end-users, carried out and being analysed at the time of writing this paper. Many of the TBD issues have been resolved or moved closer to resolution. For example, an aural component has been designed and implemented (Q2; Q3; Q4) and appears to be working better for pilots (who are accustomed to aural alerts/instructions during emergencies). Clutter (Q1) has been alleviated, and the examination of the adequacy of the visibility/salience of the green illumination in the cockpit (Q9) is under analysis.

Dual assessment mode: To accommodate different project needs, HAIQU offers two modes: Guided and User-driven. When a system is added to the app, the app asks several high-level questions about the system in the "Scoping" section. These questions investigate the AI system's maturity level (from early research to deployment-ready), EASA classification (1A to 3B), and intended capabilities (ranging from text interaction to complex gesture recognition). Then, users are prompted to choose between the Guided or the User-driven approach. If users pick the Guided approach, the app intelligently filters questions based on the answers provided in the Scoping and only shows relevant questions for their system. This step significantly cuts down the time required for the questionnaire (e.g. for FOCUS it reduced the requirements set from 180 to 120), streamlining the whole process to only relevant areas and questions. This is a quicker albeit not as deep approach. Alternatively, the User-driven approach allows users to freely explore any Human Factors area or question they deem relevant to their project, and they are free to go as in-depth as they like. Note that if the initial scoping assessment is carried out when the project is a research project, which subsequently becomes a 'real' project intended for full operational implementation, the scoping will have to be reassessed at the appropriate stage in the project. This has occurred for one of the HAIKU use cases.

Sharing for cross functional collaboration: Given the broad scope of expertise required (one person cannot answer all the questions, as they cover a range of areas or expertise), the platform is built for collaborative use by cross-functional teams, including product owners, experienced (and ideally currently licensed) end-users, AI specialists, Human Factors and safety experts. Individual users can also "share" an assessment with other users in only a few minutes, so that questions can be addressed collaboratively by users who are not in the same building, for example. That said, in its application so far (four applications), HAIQU evaluations benefit from having face-to-face sessions, as otherwise some details are often left unsaid, or else people do not always speak up when joining meetings remotely. Also, for one use case in particular, the HAIKU session spawned significant discussion on the operational concept of the Human-AI Teaming prototype itself, and the entire team found this discussion very productive, and for some of the junior members, very instructive.

Conclusions

HAIQU has already received positive feedback from the HAIKU project teams who have used it. The platform has been successfully tested with four distinct use cases: two cockpit applications, one air traffic management (ATM) application and one airport-based application. The platform has proven effective in challenging design teams to deepen their consideration of human-centric aspects of their systems. For instance, the ATM use case revealed that, while the team had addressed most requirements, they lacked robust safeguarding strategies for managing "unscripted" user interactions with the AI system. The interaction with HAIQU prompted designers to take a more detailed look at error management, making them think about edge cases they had initially overlooked.

While HAIQU has initially been developed for aviation applications, its core “formula”- translating regulatory requirements into actionable questions for design teams - has broader potential. This approach could be valuable across various safety-critical sectors, including nuclear power, oil and gas, defence and healthcare, where human-AI interaction must be carefully managed within strict regulatory frameworks.

Funding & Disclaimer

This publication is based on work performed in the HAIKU Project which has received funding from the European Union’s Horizon Europe research and innovation program, under Grant Agreement no 101075332. Any dissemination reflects the authors’ view only and the European Commission is not responsible for any use that may be made of information it contains.

References

Duchevet, A., Dong-Bach V., Peyruqueou, V., De-La-Hogue, T., Garcia, J., Causse, M. and Imbert, J-P. (2024) FOCUS: An Intelligent Startle Management Assistant for Maximizing Pilot Resilience. ICCAS 2024, Toulouse, France.

EASA (2023) EASA Concept Paper: first usable guidance for level 1 & 2 machine learning applications. February. https://www.easa.europa.eu/en/newsroom-and-events/news/easa-artificialintelligence-roadmap-20-published

EASA (2024) EASA Artificial Intelligence Concept Paper Issue 2. April. https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial-intelligenceconcept-paper-issue-2

European Parliament (2023) EU AI Act: first regulation on artificial intelligence. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-onartificial-intelligence

European Parliament (2023) EU AI Act: Recital 6. https://artificialintelligenceact.eu/recital/6/

Kaliardos, W. (2023) Enough Fluff: Returning to Meaningful Perspectives on Automation. FAA, US Department of Transportation, Washington DC. https://rosap.ntl.bts.gov/view/dot/64829

Kirwan, B. (2025: preprint) Human Factors Requirements for Aviation Human-AI Teaming. Future Transportation. doi:10.20944 https://www.preprints.org/manuscript/202501.0974/v1

Human Factors in Runway Collision: Lesson Learned from the Flight 2213 Accident Using HFACS

Mamadou Lamine Toure, Mamour Diouf, James Blundell & Wen-Chin L

Safety and Accident Investigation Centre, Cranfield University

SUMMARY

This research uses a runway collision event as a case study to analyse the complex systemic factors involved in aviation safety. The study applied the Human Factors Analysis and Classification System (HFACS) framework to analyse the Flight 2213 accident, in which an A320 Neo collided with a rescue vehicle during a take-off run. Given the involvement of both the air navigation service provider (ANSP) and the airport operator (AO) in the runway incursion, the analysis was conducted across both organisations. The findings revealed systemic failures at different HFACS levels: the ANSP exhibited deficiencies at the *Preconditions for Unsafe Acts* and *Unsafe Supervision* levels, including technological limitations, an air traffic controller's adverse physiological state, inadequate standard operating procedures, and weak supervision. The AO's failures were identified at the *Organisational Influence* and *Unsafe Supervision* levels, with a poor safety culture, ineffective coordination, and an inexperienced supervisor. The primary unsafe act was the unauthorised runway entry of the rescue vehicle. While each organisation had vulnerabilities at specific HFACS levels, their combined failures spanned all four levels, creating the conditions for the accident. These findings underscore the need for a systemic and integrated safety analysis that encompasses multiple organisation stakeholders to enhance coordination, communication, and risk management between air navigation service providers, airport operators, and airlines. Strengthening safety culture and improving inter-organisational collaboration are critical to mitigating risks and preventing future runway incursions.

KEYWORDS

Aviation Safety, Human Error, Organisational Influence, Runway Incursion, Situation Awareness

Introduction

Defined as "Any occurrence at an aerodrome involving the incorrect presence of an aircraft, vehicle, or person on the protected area of a surface designated for the landing and takeoff of aircraft." (ICAO Safety Report, 2024), runway incursions are a critical global safety priority identified by the International Civil Aviation Organization (ICAO, 2024), which significantly contributed to deaths and fatal accidents. With the continuous growth of air traffic, the likelihood of runway incursions could also increase unless enhanced safety measures are implemented. This is illustrated by the Federal Aviation Administration (FAA) 2024 report, which recorded 1,760 runway incursions in fiscal year 2023 and 1,757 incursions in 2024 (FAA, 2024).

Moreover, recent runway incursion accidents, such as those on November 22, 2022, in Lima, Peru, and on January 2, 2024, at Haneda Airport in Tokyo, serve as ongoing reminders that the risk is not yet fully controlled. In August 2024, a document titled the "Global Action Plan for the Prevention of Runway Incursions" (GAPPRI) was published as part of ICAO's Global Aviation Safety Plan and the Global Aviation Runway Safety Action Plan. Coordinated by ICAO, the Flight Safety Foundation, and Eurocontrol, and endorsed by ACI World, CANSO, and IATA, the plan was developed with the participation of over 200 experts from 80 organisations. GAPPRI is based on global and regional data analysis and offers 127 recommendations to enhance safety. These include strengthening aviation personnel training, integrating advanced technologies to improve situational awareness, optimising operational procedures and communications, and improving visual aids and infrastructure design (GAPPRI, 2024).

Significant efforts have been made to understand this phenomenon, and extensive research exists in the literature to address this multidimensional issue. A systematic review conducted by Yan, Boufous, and Molesworth (2024) highlights the key human factors influencing pilot-related runway incursions, categorising them into two main groups: failure to comply with air traffic control (ATC) instructions and positioning errors. The study underscores the importance of improving communication training, enhancing cockpit displays, and improving airport signage as key mitigation strategies (Yan et al., 2024). Another study by Hassan (2021) applied a systemic approach to analysing runway incursion risks and identified contributors including organisational and regulatory shortcomings, poorly communicated notice to airmen (NOTAMs), insufficient automation in aviation systems, and crew resource management (CRM) deficiencies. For example, an analysis of the Air Canada Flight 759 incident at San Francisco International Airport (SFO) revealed that errors in airfield lighting and pilot fatigue contributed to this near-collision event.

From a technological perspective, Omosebi et al. (2023) examine the effectiveness of various safety technologies in reducing runway incursions at U.S. airports. Their findings indicate that Runway Status Lights (RWSL) significantly reduce severe incursions (Categories A and B) by providing immediate and direct alerts to pilots. In contrast, the Airport Surface Detection Equipment Model X (ASDE-X) is deemed less effective due to its reliance on ATC intervention, which can delay hazard recognition. The study highlights the need to prioritise RWSL deployment at high-risk airports with complex runway configurations, along with strategic improvements in airport design to minimise runway intersections (Omosebi et al., 2023).

One of the most widely adopted frameworks for comprehensive accident analyses is the Human Factors Analysis and Classification System (HFACS). Developed by Wiegmann & Shappell (2001), HFACS builds on the Swiss Cheese Model of Reason (1990) by categorising accident causation across four hierarchical systemic levels: unsafe acts, preconditions for unsafe acts, unsafe supervision, and organisational influences. HFACS provides a structured taxonomy that classifies both active failures of operators and latent conditions across the aviation system. Its application has been extended beyond aviation to industries such as maritime, rail, and mining (2019), demonstrating its versatility and effectiveness in accident causation analysis.

The literature surrounding HFACS highlights its effectiveness in investigating accidents or incidents at the various levels of an organisation. Mohandas & Weng (2021) applied HFACS to 75 aviation accidents in Singapore, identifying critical human errors and organisational shortcomings that aligned with official investigation reports. Their findings underscored HFACS's value in classifying both immediate causes, such as skill-based errors, and latent factors, such as training deficiencies and inadequate supervision. Similarly, Small (2020) applied HFACS to the Asiana Airlines Flight 214 crash, revealing how poor pilot training, organisational oversight, and procedural violations collectively contributed to the accident. The study concluded that focusing solely on pilot errors without addressing broader organisational influences leads to incomplete safety interventions. Also, Yan et al. (2025) conducted a HFACS analysis of pilot-related runway incursions in the United States and Australia, identifying teamwork failures and communication breakdowns as the most prominent contributors to runway incursions; *Failure of effectively communicate* and *inadequate communication equipment* being the most frequently observed factors. Also, procedural violations played a significant role across multiple phases of events. At higher system levels, *organisational policy risks not adequately assessed* was the most frequently cited.

Despite HFACS's demonstrated effectiveness, it is found that its 'unit of analysis' is commonly limited to single organisation. This study will apply the framework across two relevant organisations within the analysis of the LATAM Flight 2213 accident to identify the between and within organisational factors, and their interactive dynamics, that contributed to the event. By systematically categorising both active and latent failures, the study seeks to offer insights into pilot behavior, air traffic controller communication, airport infrastructure, and organisational practices. Through this comprehensive approach, the study will provide more robust evidence-based recommendations for improving safety in aviation operations and reducing the occurrence of future incidents or accidents, contributing to the ongoing enhancement of aviation safety management systems.

Methods

The Flight 2213 accident was analysed using HFACS as a case study to identify the causal factors contributing to runway incursions. The accident report of Perú Flight 2213 contains detailed information about the A320 Neo aircraft collision with a firefighter vehicle that was crossing runway 16L at Jorge Chávez International Airport on 18 November 2022. HFACS is a widely used framework, including four levels and eighteen categories related to human factors in flight operations (Wiegmann & Shappell, 2003). This study applied the HFACS framework across the two organisations involved in the accident, the Air Navigation Service Provider (ANSP) and the Airport Operator (AO).

In this work, an air traffic controller and an air traffic manager (both with eight years of experience), as well as two chartered experts in human factors, used HFACS to analyse the accident report of the runway incursion and developed accident prevention strategies.

Results and Discussion

The accident was primarily caused by an unsafe act committed by the driver of rescue vehicle, who entered the runway without authorisation from the ANSP air traffic controller (ATC). The investigation identified two preconditions for unsafe acts within the air navigation service provider organisation: technological environmental factors and adverse physiological state.

Regarding the technological factors, it was found that ATCs at the ANSP tower lacked access to technology that could enhance their situational awareness because they only had binoculars to locate the position of vehicles, which was insufficient to allow a better understand the situation (i.e., weakness of level 1 situation awareness). Additionally, the analysis revealed that the surface controller was undergoing rehabilitation from a broken leg, placing him in an adverse physiological state. This situation would have prevented him from standing up to follow the traffic taking off as part of the ATCO task. These preconditions likely contributed to the occurrence of the accident.

The analysis further indicated that the accident was influenced by deficiencies at the *Organisational Process* level. Specifically, there was an inadequate SOP detailing the duration of the emergency vehicle's response time (the Exercise Time Response (ETR)), resulting in a lack of coordination and communication between the AO and ANSP. Furthermore, there was a lack of adequate procedures regarding the shift change process at the ANSP tower because the outgoing and upcoming shift supervisor did not provide information to the joining controllers regarding the execution of the second ETR exercise.

At the *Unsafe Supervision* level, two contributing factors were identified: inadequate supervision and planned inappropriate operations. There was a lack of supervision within the ANSP tower at the time of the accident and an absence of coordination with the AO regarding the exercise. Notably, there was no unsafe act committed by an operator within the ANSP organisation itself.

For the AO, the investigation revealed two issues at the *Organisational Influence* level: organisational process and organisational climate. The findings indicated a poor safety culture within the AO and an inadequate SOP for the ETR exercise, particularly concerning coordination, and communication. At the *Unsafe Supervision* level, there was a lack of coordination within AO, and the supervisor of the ETR exercise was found to be inexperienced. The identified unsafe act was the unauthorised entrance of the rescue vehicle onto the runway without ATC clearance. Unlike the ANSP, no preconditions for unsafe acts were identified within the AO.

Applying the HFACS model to both organisations involved in the accident – the ANSP and AO – it was revealed that a complementary distribution of failures existed across the four levels of the framework. A summary of this analysis is presented in **Table 1**.

Individually, each organisation exhibited weaknesses at specific HFACS levels while remaining free of failures at others. The ANSP was primarily affected at the *Preconditions for Unsafe Acts* level and the *Unsafe Supervision* level. On the other hand, the AO showed deficiencies at the *Organisational Influence* level and the *Unsafe Supervision* level.

However, when both analyses are combined, failures are present across all four HFACS levels**.** This comprehensive perspective highlights the interdependence of organisational failures. While each entity has its vulnerabilities, their interaction creates an environment conducive to accidents, where every weak link contributes to the outcome. This analysis underscores the need for a systemic and integrated approach to safety, involving both air navigation service providers and airport operators, to address gaps at all levels and prevent future accidents.

Table 1: Summary of findings

HFACS Level	**ANSP**	**AO**	**Combined Organisation**
ORGANISATIONAL INFLUENCE	🟧 Inadequate SOPs: Poor shift change procedures 🟧 Lack of Coordination: Poor communication with the AO	🟧 Poor Safety Culture: Weak internal safety climate 🟧 Inadequate SOPs: Poor coordination and communication for ETR exercise	🟧 Failures at this level in both organisations, but different aspects
UNSAFE SUPERVISION	🟧 Inadequate Supervision: Lack of supervision within the ATC tower 🟧 Planned Inappropriate Operations: No coordination with the AO on the exercise	🟧 Inadequate Supervision: Lack of internal coordination 🟧 Supervisor Inexperience: The ETR exercise supervisor lacked experience	🟧 Failures in unsafe supervision exist in both organisations
PRECONDITIONS FOR UNSAFE ACTS	🟧 Technological Environmental Factor: Lack of ATC situational awareness tools 🟧 Adverse Physiological State: Surface controller recovering from injury	🟩 No identified preconditions for unsafe acts	🟧 Present in ANSP and absent in AO

UNSAFE ACTS	No unsafe acts identified	Unauthorised runway entry by rescue vehicle	Unsafe act present in AO but not ANSP
			ACCIDENT

Color Lexicon

- **Green**: No failure detected at this HFACS level.
- **Orange**: Failure detected at this HFACS level.
- **Red**: The accident occurred due to cumulative failures at four levels.

Conclusion

The runway incursion involving a rescue vehicle was a preventable event resulting from a combination of systemic organisational deficiencies, including inadequate supervision, ineffective communication protocols, insufficient safety culture, and flawed procedural controls. This analysis has highlighted that aviation accidents rarely stem from isolated operator errors but are instead the product of multiple interrelated latent failures across organiational levels. Additionally, this study revealed that while individual organisations exhibit weaknesses at specific HFACS levels, their combined failures span across all four levels, emphasising the interconnected nature of organisational deficiencies. HFACS can provide a comprehensive framework for identifying and mitigating these latent risks, demonstrating its efficacy in aviation safety management.

ICAO has published the GAPPRI, which highlights variability in human performance and miscommunication as leading factors in runway incursions. These findings align closely with the results of this case study. The Flight 2213 accident resulted from systemic failures at multiple levels, including poor organisational coordination, supervisory lapses, environmental complexities, and individual unsafe acts leading to human factors accidents/incidents. The chain of events was initiated by a lack of joint planning and culminated in critical communication breakdowns and procedural violations. Future research may develop artificial intelligence to detect unauthorised vehicle/aircraft movements near active runways and issue timely alerts to air traffic controllers and pilots simultaneously.

To prevent runway incursions and enhance safety, effective coordination between ANSPs, AOs, and emergency teams should be reinforced, with emergency drills scheduled during low-traffic periods. Advanced technology, including real-time runway monitoring and automated clearance systems, must be integrated to improve situational awareness. A strong safety culture should be fostered through transparent reporting mechanisms within a well-structured Safety Management System (SMS). Finally, strict compliance with ICAO regulations, particularly Annex 14 and Annex 19, along with regular safety audits, is crucial to maintaining operational efficiency and mitigating risks.

By implementing these measures, aviation stakeholders can address safety gaps across all four HFACS levels, creating a resilient safety ecosystem that minimises the likelihood of runway incursions and enhances overall operational efficiency.

However, this study is not without limitations. While HFACS is a robust framework, its effectiveness is dependent on the quality and availability of data collected during accident investigations. Incomplete or inaccurate data may lead to an incomplete identification of causal factors, particularly at the higher organisational levels. The integration of HFACS with other safety frameworks, such as Systems-Theoretic Process Analysis (STPA)**,** may enhance the system-wide identification of hazards and provide a more holistic view of aviation safety for runway collision events.

References

Federal Aviation Administration. (2024, November 28). *Runway safety statistics*.

Harris, D., & Li, W. (2019). Using neural networks to predict HFACS unsafe acts from the preconditions of unsafe acts. *Ergonomics, 62*(2), 181-191. https://doi.org/10.1080/00140139.2017.1407441

Hassan, M. (2021). *An investigation of runway risks using a systems approach* (Master's thesis). Florida Institute of Technology, Melbourne, Florida. Retrieved from https://repository.fit.edu/etd

Hulme, A., Stanton, N. A., Walker, G. H., Waterson, P., & Salmon, P. M. (2019). Accident analysis in practice: A review of Human Factors Analysis and Classification System (HFACS) applications in the peer-reviewed academic literature. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 63*(1), 1849–1853. https://doi.org/10.1177/1071181319631086

International Civil Aviation Organization. (2024). *Safety report: 2024 edition*. International Civil Aviation Organization. https://www.icao.int/safety/Documents/ICAO_SR_2024.pdf

International Civil Aviation Organization, Flight Safety Foundation, EUROCONTROL, Airports Council International, Civil Air Navigation Services Organization, & International Air Transport Association. (2024). *Global action plan for the prevention of runway incursions (GAPPRI)*.

Mohandas, H. R., & Weng, T. K. (2021). Human factors analysis for aviation accidents and incidents in Singapore. In D. Harris & W.-C. Li (Eds.), *Engineering psychology and cognitive ergonomics* (pp. 308–323). Springer International Publishing.

Reason, J. (1990). *Human error*. Cambridge University Press.

Reason, J. (1997). *Managing the risks of organizational accidents*. Ashgate Publishing.

Small, A. (2020). Human Factors Analysis and Classification System (HFACS): As applied to Asiana Airlines Flight 214. *The Journal of Purdue Undergraduate Research, 10*(1). https://doi.org/10.7771/2158-4052.1485

Wiegmann, D. A., & Shappell, S. A. (2001). A human error analysis of commercial aviation accidents using the Human Factors Analysis and Classification System (HFACS).

Wiegmann, D. A., & Shappell, S. A. (2003). *A human error approach to aviation accident analysis: The Human Factors Analysis and Classification System (2003rd, 2017th ed.)*. Ashgate Publishing.

W.T.-K. Chan et al. (2024). Evaluating cause-effect relationships in accident investigation using HFACS-DEMATEL. Retrieved from https://dspace.lib.cranfield.ac.uk/server/api/core/bitstreams/d9b7802e-862e-4130-9600-c209aab8170e/content

Yan, Y., Boufous, S., & Molesworth, B. R. C. (2024). A systematic review of pilot-related runway incursions from a human factors perspective. *The International Journal of Aerospace Psychology, 34*(1), 1-19.

Yan, Y., Boufous, S., & Molesworth, B. R. C. (2024). Speaking of human factors: An interview study on the causes and prevention of runway incursions with aviation professionals. *Preprint, SSRN*. Available at: https://ssrn.com/abstract=4938916

Yan, Y., Boufous, S., & Molesworth, B. R. C. (2025). Pilot-related factors involved in runway incursions: An epidemiological approach. *The International Journal of Aerospace Psychology, 1*-20. https://doi.org/10.1080/24721840.2025.2452443

Human Performance Optimisation Interventions and Measures of Effectiveness in Air Traffic Control

Katie Fisher & Leigh-Anne Smith

NATS

SUMMARY

This paper outlines three programmes forming the Human Performance Optimisation (HPO) area of work within the Human Factors (HF) and Human Performance (HP) team at NATS, Air Traffic Control (ATC): HP Coaching, Performance Mindset and the HP Buddy Programme. Pre-cursors and drivers, development, application and analysis are described as well as additional work to evaluate and improve effectiveness measures for single and combined interventions.

KEYWORDS

Human Performance Optimisation (HPO), Coaching, Air Traffic Control (ATC)

Introduction

Training and working as an ATC Officer (ATCO) is rewarding but can bring with it operational and personal challenges, as well as perceived pressures that have the potential to impact on workplace performance. For example, loss of confidence may be experienced by a trainee after a challenging training session or by an ATCO after a return to work or post incident.

With safety the key priority, it is well recognised that viewing non-technical aspects of performance alongside the technical can support the optimisation of performance as well as contribute positively to individual experiences of work and personal well-being (Flin et al, 2008).

The HP Optimisation Work Stream

The benefits of optimising human performance in NATS as a safety critical industry are wide ranging and include improving understanding, management and awareness of non-technical factors that underpin and enhance technical skills. Non-technical factors include teamwork, communication, information processing (decision making, situational awareness, workload and threat and error management), human resilience (including confidence, adaptability, purpose, and social support; stress and fatigue management). Application of these skills alongside the technical combine to support the improvement of safety and drive positive behaviour and culture.

In recognition of the above, the HPO workstream at NATS consists of three inter-related areas which have evolved over six years. The programme comprises interventions designed specifically to support, develop and optimise performance through the provision of one- to- one coaching time, facilitated discussion-based sessions and the sharing of relevant tools and resources.

The HP Coaching Programme

The benefits of coaching to support individual performance and development are widely recognised. The HP Coaching Programme was set up in 2019 to make strengths based coaching support available for all operational roles within the ATC environment to help overcome non-technical challenges. The solutions focussed OSKAR model (Jackson and McKergow, 2002) has been the preferred approach taken due to a forward focus and has been applied consistently across coaching sessions to support the identification of outcomes and actions to develop non-technical awareness and skills. Six structured coaching sessions are offered as the optimal number to provide an opportunity to establish new habits and work towards specific goals (Rock and Schwartz, 2014).

Coaching is only of benefit when individuals are open to the challenge of self -reflection and development (Cox et al, 2018). HP Coaching is therefore not mandatory or provided to all operational trainees but is available and allocated on request, often in response to a training or performance- based challenge. A referral is made via a line or training manager in collaboration with the individual.

During 2024, the HP Coaching team provided 120 coaching sessions with key session themes being confidence (65%), mindset (50%), management of perceived pressure and nerves (33%) and management of feedback (21%).

The Performance Mindset Programme

The Performance Mindset group intervention was introduced in 2023 to address the most common HP challenges faced in ATC training. It was recognised that whilst HP Coaching support was able to address challenges for individuals on a one to- one basis, the knowledge, tools and strategies discussed within coaching sessions could benefit all trainees if packaged into a group intervention as a proactive method of support.

Programme content was developed following thematic analysis of topics arising in HP Coaching sessions with ATC trainees in 2021/22. Mindset was identified to be the key theme within non-technical skills. Other areas included assertiveness, confidence, perceived pressure, emotional resilience, motivation and impact of emotions on performance.

Collaborating with Unit and Training Managers, themes emerging from coaching data at distinct phases of training and pinch points in the training pathway were identified alongside additional relevant content. The Performance Mindset programme was developed to consist of five core sessions spaced strategically across the training pathway from basic training through to near validation. It comprises of facilitated sessions to introduce learners to a range of research-based techniques around non-technical skills applicable to the ATC context. Sessions provide the space for learners to share experiences, discuss strategies for overcoming challenges and use resources and tools for personal development. Session content includes topics such as values, strengths, healthy performance habits, goal setting, building confidence and locus of control. Since the inception of the programme, 71 sessions have been delivered with 272 learners having attended at least one session.

Performance Mindset sessions reinforce principles and learning covered by regulatory content such as those relating to stress and error management, fatigue, communication and well-being. Over time and after review of feedback and reported benefits, sessions are mandatory and timetabled throughout the training programme.

The Human Performance Buddy Programme

The HP Buddy Programme was set up in early 2024 to combine coaching for non-technical skills with a mentoring approach. Key drivers for implementation were to improve sharing of knowledge, skills and experience already present in the organisation and as a solution for the high demand for HP Coaching. The programme provides first line support to learners in ATC and engineering in both technical and non-technical skills and has been achieved by training HP Buddies to work with learners in the same role to facilitate the improvement of performance and help navigate through training challenges. Benefits to the business include increased support for learners, earlier and more accessible intervention and efficient use of resource, optimising the knowledge, skill and experience available.

HP Buddies receive training in effective coaching skills and use of HP resources in non-technical areas relevant to the training environment and take an approach that supports and encourages learners to be self-led and proactive in their own development. The programme uses the GROW model of coaching (Whitmore, 1992) as the preferred approach for simplicity and goal focus. An additional key role of an HP Buddy is to triage, signpost and refer to alternative support if required.

There are currently 20 trained HP Buddies across the organisation in both ATC and engineering roles. 17 individuals have received support over 29 sessions, key themes being confidence, resilience, visual scanning, ATC planning tasks and mindset.

Feedback, Data Collection and Analysis

For each intervention, data is continuously collected and used to make iterative improvements across the programme. Number of referrals, sessions delivered, themes of support, topics discussed, feedback from referrers, coaches and delegates is recorded for analysis and review. Subjective feedback is also gathered to indicate the benefits of the interventions in support of goals (for example passing a phase check or gaining validation as an ATCO:

"I had my validation board yesterday and was successful. Thank you for all your help, I had a lot of confidence going into the board after our sessions. Many thanks" (Coachee)

"You'll be pleased to hear I had a really good day of controlling yesterday. After one session in particular I got a very positive debrief full of praise, within earshot of at least 5 members of my OJTI team. That felt really, really good. You'll say now that it's all because of my hard work, but it's also because of your help, so really thank you very much!" Coachee)

Hopefully the second session was as positive and constructive as the first……X has definitely been demonstrating a more positive mindset since your first session together. Cheers. (Referrer)

This type of feedback highlights the coaching support provided is perceived to provide a positive benefit by both learners and referrers in terms of improved experience, performance and results.

Future work may involve widening the scope to include On the Job Training Instructors (OJTIs), Operational ATCOs and making content more accessible such as extending Performance Mindset sessions to interactive online content.

Evaluating the Effectiveness of Human Performance Optimisation Interventions

Although significant subjective evidence of the quality of these programmes has been reviewed, a lack of objective data hindered efforts to demonstrate return on investment (ROI), refine interventions, and secure resources. With rising engagement, it is critical to evaluate the effectiveness of these programs to sustain and optimise their contribution to training in a safety-critical environment.

Method

The project employed the Define, Measure, Analyse, Improve and Control (DMAIC) framework, a structured approach to process improvement. Lean and Six Sigma tools, including process analysis, root cause analysis, and improvement analysis were used to evaluate current processes and identify opportunities for improvement. The project aimed to understand the effectiveness of HPO interventions since the introduction of Performance Mindset in May 2023, to define the benefits of both independent and combined interventions.

Outcomes

Define

The problem statement was defined as: *"How can we track the effectiveness and benefits of structured human performance interventions at NATS?"* The extent of the problem was considered, identifying key challenges in assigning budgets due to a lack of demonstrable return on investment, targeted improvement of interventions, stakeholder engagement and reduced job satisfaction.

A Supplier, Inputs, Process, Output, Customers (SIPOC) analysis revealed the existing process to be error prone and labour intensive for practitioners.

Measure

Qualitative and quantitative data were collected from stakeholders and practitioners regarding their experience of gathering, analysing, and presenting data on HPO intervention effectiveness through a series of interviews and surveys.

A database was created to map data collected pertaining to HPO interventions to effectiveness outcomes, enabling identification of key wastes. Reviews of quantitative data demonstrated the cost of intervention provision to date highlighting significant investment with unproven effectiveness.

Analyse

An initial brainstorming tool identified emerging causes of challenges in the process of evaluating HPO intervention effectiveness according to "people", "process" and "technology" from the data collected in the Measure phase. An affinity diagram was completed resulting in seven emergent themes. A root cause analysis using the "five whys" method (Liker, 2004) identified the primary issue as: *"Definitions of effective HPO interventions have never been clearly defined."*

Due to the nature of the problem presented, it was identified that outlining a definition of effectiveness would be insufficient to define the benefits brought about by HPO interventions. Additional causes were considered from the affinity diagram and barriers to improve were agreed: Data is highly dispersed; Data quality is poor (fragmented and incomplete datasets); Data analysis processes are manual and time consuming.

Improve

An initial improvement tool to simplify, eliminate, combine, automate or reallocate (SECAR) steps of the existing process was completed against each key root cause. An Ease-Benefits tool identified appropriate improvement steps to define, measure and present data.

Agreed steps included: Generate definitions of effectiveness for each intervention; Update data gathering methods to align data points to effectiveness marker; Align data collected across the training organisation to effectiveness markers; Create dashboard to demonstrate HPO intervention contributions to training outcomes.

To create the dashboard, a range of effectiveness statements were mapped to quantitative data gathered through feedback forms and HP Coaching Session Records (Table 1).

Table 1: Effectiveness Statements with data mapping to data held within Human Performance

HPO Intervention	Effectiveness Statement	Data Location	Scale of Measurement
HP Coaching	There are observable changes in Confidence, Resilience and Response to Feedback	Coachee and Referrer Feedback Forms	Nominal
HP Coaching	HP Coaching is recommended	Coachee and Referrer Feedback Forms	Interval
HP Coaching	HP Coaching improves self-perceived performance in problem specific areas.	Coachee Feedback Forms, HP Coaching Session Record	Interval
Performance Mindset	Learners embed learning from sessions.	Performance Mindset Feedback Forms	Nominal
Performance Mindset	Performance Mindset is recommended	Performance Mindset Feedback Forms	Interval
Performance Mindset	Performance Mindset improved self-rated Confidence, Resilience and Response to Feedback	Performance Mindset Feedback Forms	Nominal

Additional quantitative measures were included for the effectiveness statements in Table 2. Correlations between scores on training objectives over the time HP Coaching was received by an individual were completed to demonstrate impact of HP Coaching on overall and individual objective achievement. Training data was also correlated with the number of HP Coaching sessions received by an individual to demonstrate optimum length of engagement in HP Coaching.

Table 2: Effectiveness statements with data mapping to data held in Operations Training records

HPO Intervention	Effectiveness Statement	Data Location	Scale of Measurement
HP Coaching	HP Coaching improves performance in assessment of ATC specific skills	TRACER Individual Training Record	Nominal
HP Coaching	HP Coaching reduces the need for instructor guidance in specific ATC skills	TRACER Individual Training Record	Nominal
HP Coaching	There is an optimum number of HP Coaching sessions for improved ATC performance	TRACER Individual Training Record, HP Coaching Session Record	Ratio
Performance Mindset	Completion of the full Performance Mindset programme is associated with reduced overall training duration	Performance Mindset Tracker, Average Training Time	Ratio

Control

The implementation of a dashboard as a minimum viable product is underway by working collaboratively with Operations Training and Analytics teams. The preliminary version of the dashboard provided a centralised view of intervention effectiveness and has undergone initial changes following stakeholder feedback, including the display of data for easier customer readability and additional measures of effectiveness. For example, an updated version of the dashboard will include correlation analysis of topics covered by individuals in the Performance Mindset sessions and those explored in HP Coaching, to understand whether group interventions mitigate the requirement for one- to- one support on the same topic at the same level in ATC training.

Moving forward, the dashboard will allow practitioners and stakeholders to monitor performance of the programme, identify trends, and refine strategies as needed. Ongoing monitoring and updates will ensure the process remains robust and adaptive to future challenges. Planned controls to identify continuous improvement actions include stakeholder surveys and practitioner feedback on the process of data provision for the model. Future improvements to the dashboard have been identified throughout the process, requiring further data collection and the inclusion of HP Buddy data. In the future, the dashboard will demonstrate the impact of HPO interventions on predicted training duration for ATC trainees.

Conclusion

The HPO interventions at NATS provide a layered approach to support including coaching skills in conversations, human performance coaching and group support to develop and learn tools, strategies and share experiences. This combination of support enables individuals to improve their non- technical skills alongside the technical to optimise progress. The layered approach has a number of advantages, i.e. early triaging of HP issues, increased awareness of support available, proactive as well as reactionary measures and an expanded reach compared to a single mechanism. This model could be applied to other settings outside of ATC where resources are tailored specifically to the environment.

HP interventions can be measured and communicated using a structured approach. Defining intervention-specific metrics and aligning data collection processes with these measures are critical to ensuring transparency and optimizing outcomes. Dashboards offer a practical solution for presenting complex data in an accessible way.

By adopting an improvement methodology, NATS can enhance the credibility and sustainability of its HPO interventions, ensuring continued support for non-technical skills training in a safety-critical environment. These findings contribute to a growing body of evidence supporting the need for rigorous evaluation frameworks in human performance initiatives. Widening the support to other areas of the business such as On the Job Training Instructors and making content more accessible such as extending Performance Mindset sessions to interactive online content are all in the pipeline.

References

Cox, E., Bachkirova, T., & Clutterbuck, D. (2018). *The Complete Handbook of Coaching* (4th ed.). Sage Publications.

Flin, R., O'Connor, P., & Crichton, M. (2008). *Safety at the Sharp End: A Guide to Non-Technical Skills*. Ashgate Publishing.

Jackson, P. Z., & McKergow, M. (2002). *The Solutions Focus: The SIMPLE Way to Positive Change*. Nicholas Brealey Publishing.

Liker, J. K., *The Toyota Way: 14 Management Principles from the World's Greatest Manufacturer*, McGraw-Hill Education, 2004.Rock, D., & Schwartz, J. (2006). The neuroscience of leadership. *Strategy and Business*, 43.

Whitmore, J. (1992). *Coaching for Performance: GROWing Human Potential and Purpose: The Principles and Practice of Coaching and Leadership*. Nicholas Brealey Publishing.

Impact of Aircraft Synoptic Page Designs on Pilot Flight Performance

Kübra Bager, Gamze Sevimli, Cemre Aymelek, Gökhan Bayramoğlu, Canan Angın, Gizem Bodur Uruç & Atakan Coşkun

Turkish Aerospace

SUMMARY

This study evaluates the impact of a newly designed Hydraulic Synoptic page on pilot performance during emergency scenarios in simulated flights. By comparing it with the traditional Hydraulic page, the research assesses pilot workload, situational awareness, and system usability. Results suggest that the synoptic page significantly enhances performance, safety, and usability, offering valuable insights for cockpit interface design.

KEYWORDS

Aerospace, Situational Awareness, Usability

Introduction

This study examines the impact of a newly designed Hydraulic Synoptic page on pilot flight performance under emergency scenarios during simulated flight operations. The analysis focuses on pilot workload, situational awareness, and the usability of the synoptic page in comparison to the traditional Hydraulic page. These factors are crucial in ensuring operational efficiency and safety in aviation, particularly in critical scenarios.

Workload can be defined as the integrated mental and physical effort required to satisfy the perceived demands of a specified task. In the aviation concept, pilot workload can be defined as the demand placed on the pilot's mental and physical resources. During the flight operation, mental processes such as attention, perception, decision-making, and also physical cockpit controls are the factors that affect workload (Hicks, Durbin, Morris, & Davis, 2014).

Situational Awareness (SA) can be defined as the pilot's perception of the elements of the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future. Pilots must know the state of their aircraft, the environment through which they are flying, and relationships between them (Endsley & Robertson, 2000).

Usability is a quality attribute that assesses how easy user interfaces are to use. In addition, usability depends on how well design's features accommodate users' needs and contexts. Satisfaction, learnability, efficiency, few errors and memorability will be the main focus while measuring the usability of a system (Laubheimer, 2018).

Methodology

Five test pilots participated in two randomized simulated flight scenarios, each designed to simulate distinct hydraulic failure alerts:

- Scenario A: Utilised the traditional Hydraulic Page
- Scenario B: Employed the new Hydraulic Synoptic page.

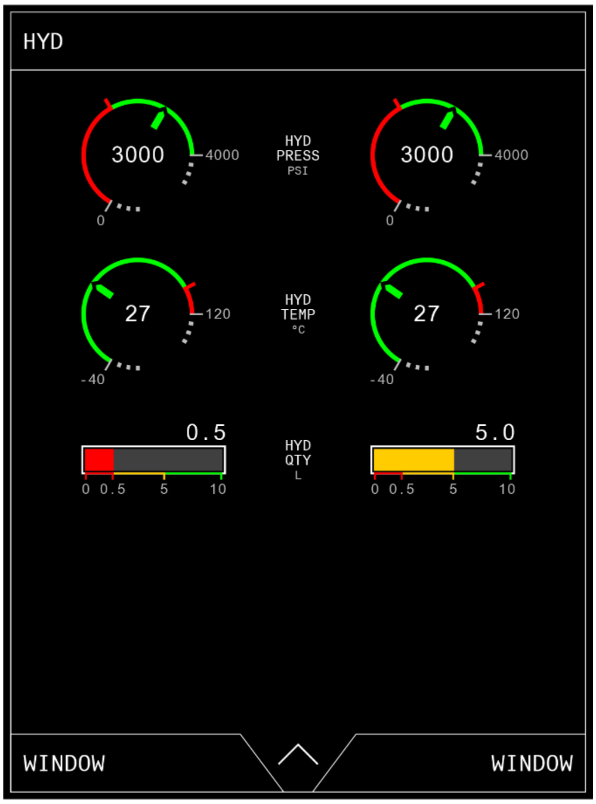


Figure 1: Traditional HYD Page

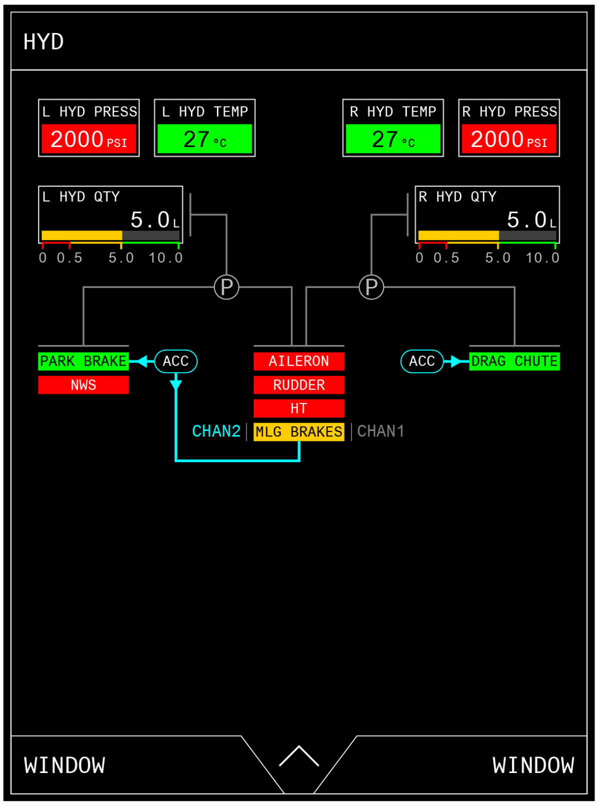


Figure 2: HYD Synoptic Page

During both scenarios, pilots conducted take-off, cruise, and landing phases, encountering hydraulic failure alerts during landing. The names of alerts were Left Hydraulic Quantity (L HYD QTY) and Left Hydraulic Pressure (L HYD PRESS). They performed predefined alert procedures while interacting with the respective Hydraulic page. Workload ratings (Bedford Scale), situational awareness scores (SART), and usability evaluations (SUS) were collected. Flight data such as response time and landing success were recorded.

Table 1: Alert Procedures

L HYD QTY	L HYD PRESS
1) LAND AS SOON AS PRACTICAL	1) LAND AS SOON AS PRACTICAL
WHEN TWO-POINT TOUCH DOWN; 2) DRAG CHUTE - DEPLOY 3) USE BRAKES TO DECELERATE A/C	WHEN TWO-POINT TOUCH DOWN; 2) DRAG CHUTE – DEPLOY 3) USE BRAKES TO DECELERATE A/C
4) DO NOT ENGAGE NWS	4) DO NOT ENGAGE NWS
5) USE DIFF BRAKING INSTEAD OF NWS	5) USE DIFF BRAKING INSTEAD OF NWS
6) USE BRAKE PEDALS TO HOLD A/C	6) USE BRAKE PEDALS TO HOLD A/C
7) PARK BRAKE SWITCH - PARK BRAKE	7) PARK BRAKE SWITCH - PARK BRAKE

Table 2: Details of Test Scenario

Scenario	Alert	Instrumentation	Independent Variables	Dependent Variables
Scenario A	L HYD QTY	Bedford Rating Scale Situational Awareness Rating Scale Qualitative Form Eye tracking data System Usability Scale	Hydraulic Page Design – Hydraulic failure situation without synoptic page	1. Situational awareness of the pilot 2. Pilot's workload 3. Simulator flight data (e.g., alert procedure completion & time spent, flight success, landing phase duration, pilot reactions to cockpit controls)
Scenario B	L HYD PRESS	Bedford Rating Scale Situational Awareness Rating Scale Qualitative Form Eye tracking data System Usability Scale	Hydraulic Synoptic Page Design – Hydraulic failure situation with synoptic page	

Instrumentation

1. Bedford Rating Scale (BRS)

After completing each task scenario, participants will be asked to complete the Bedford Rating Scale items (Figure 3).

Workload Success Criteria

The Bedford Rating Scale results will be evaluated according to the criteria as shown in Figure 4.

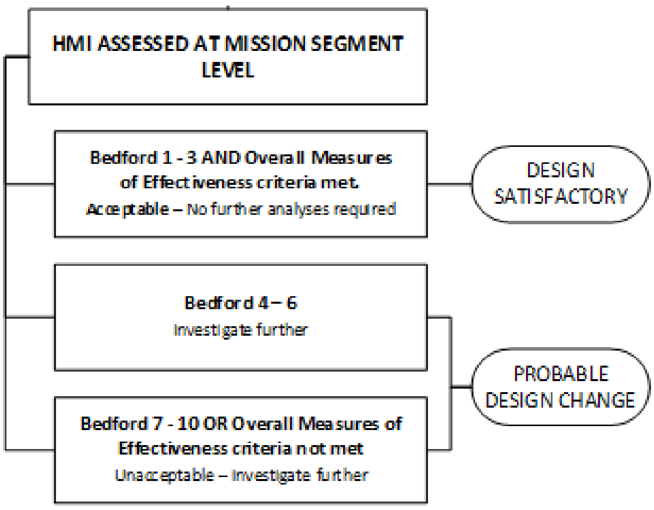


Figure 3: Bedford Rating Scale Interpretation

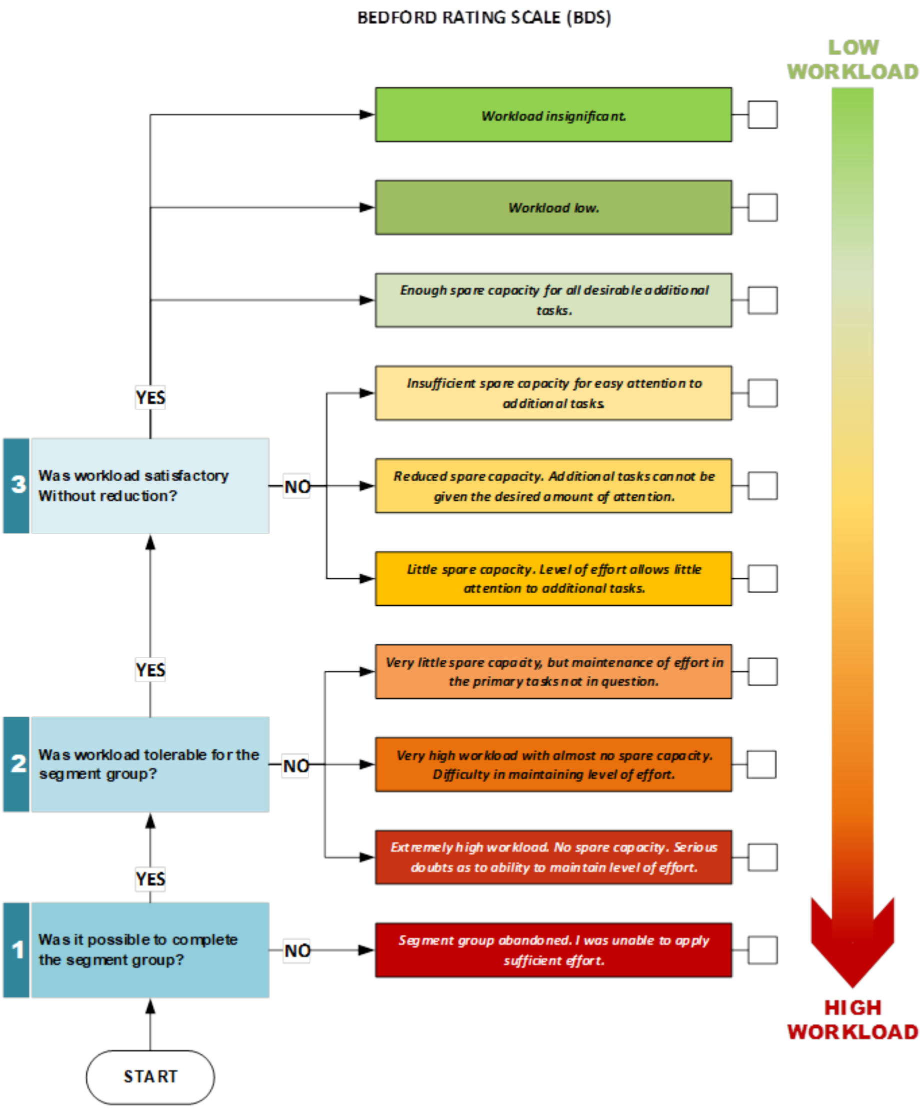


Figure 4: Bedford Rating Scale

2. Situational Awareness Assessment Method

Situational Awareness Rating (SART) will be used to measure SA (Tayor, 1990). SART technique allows subjective estimation of SA. It includes 10 dimensions which are used in conjunction with a likert scale. After completing each task scenario, participants will be asked to rate the SART rating scale items as given below.

Table 3: SART Rating Sheet

Plight Phase	Query No								
Demand	1. How changeable is the situation?								
	Stable and straightforward	1	2	3	4	5	6	7	Changing suddenly
	2. How many variables are changing within the situation?								

	Very few variables changing	1	2	3	4	5	6	7	A large number of factors varying
	3. How complicated is the situation?								
	Simple and straightforward	1	2	3	4	5	6	7	Complex with many interrelated components
Supply	**4. How alert are you in the situation?**								
	A low degree of alertness	1	2	3	4	5	6	7	Alert and ready for activity
	5. How much mental capacity do you have to spare in the situation?								
	Nothing to spare at all	1	2	3	4	5	6	7	Sufficient to attend to many variables
	6. How much are you concentrating on the situation?								
	Focusing on only one	1	2	3	4	5	6	7	Concentrating on many aspects of the situation
	7. How much is your attention divided in the situation?								
	Focusing on only one	1	2	3	4	5	6	7	Concentrating on many aspects of the situation
Understanding	**8. How much information have you gained about the situation?**								
	Very little	1	2	3	4	5	6	7	A great deal of knowledge
	9. How good is the information you have gained about the situation?								
	It is not usable at all	1	2	3	4	5	6	7	The knowledge communicated very useful
	10. How familiar are you with the situation?								
	It is a new situation	1	2	3	4	5	6	7	A great deal of relevant experience

Situational Awareness Success Criteria

The higher score indicates a better level of SA, while a lower score suggests a lower level of SA. Minimum SART rating score is -14, while maximum SART rating score is 46. Therefore, it can be said that if SART rating is greater than 16, situational awareness level will be considered as acceptable.

3. Usability Assessment Scale

To measure usability, (SUS) technique will be used (Laubheimer, 2018). After completing each task scenario, participants will be asked to rate the usability scale items (Table 4).

Table 4: System Usability Scale (SUS)

Item #	Test Item	Strongly Disagree 1	2	3	4	Strongly Agree 5
1	I would like to use this system frequently.	◯	◯	◯	◯	◯
2	I found the system unnecessarily complex.	◯	◯	◯	◯	◯
3	The system was easy to use.	◯	◯	◯	◯	◯
4	I would need the support of a technical person to be able to use this system.	◯	◯	◯	◯	◯
5	I found the various functions of this system were well integrated.	◯	◯	◯	◯	◯
6	I thought there was too much inconsistency in this system.	◯	◯	◯	◯	◯
7	I would imagine that the most pilot would learn to use this system very quickly.	◯	◯	◯	◯	◯
8	I found the system very cumbersome to use.	◯	◯	◯	◯	◯
9	I felt very confident using the system.	◯	◯	◯	◯	◯
10	I need to learn a lot of things before I could get going with this system.	◯	◯	◯	◯	◯

Findings

The results reveal a significant impact of the Hydraulic Synoptic page on pilot workload, situational awareness, and usability:

1. **Workload Assessment (Bedford Scale):**

   Both Scenario A (Traditional Hydraulic Page) and Scenario B (Hydraulic Synoptic Page) received a workload rating of 1, indicating that pilots experienced minimal effort in managing the hydraulic failure alerts. This suggests that the redesigned page does not add cognitive burden, ensuring that pilots can efficiently handle emergency situations without increased stress or fatigue.

2. **Usability Evaluation (System Usability Scale - SUS):**

   The usability of both interfaces was rated highly, with Scenario A scoring 95 and Scenario B achieving a slightly higher 97.5. The improved score in Scenario B reflects better intuitiveness, ease of navigation, and pilot confidence in interacting with the synoptic display. Pilots reported that the synoptic design facilitated quicker identification of hydraulic system status and necessary actions.

3. **Situational Awareness (SART Scores):**

   Pilots using the Hydraulic Synoptic Page (Scenario B) achieved a SART score of 22, compared to 19 in Scenario A. The higher score indicates improved perception, comprehension, and projection of the system's operational state. The graphical representation and structured information layout of the synoptic page contributed to faster decision-making and a more accurate understanding of system behavior.

4. **Pilot Feedback and Observations:**

   Pilots noted that the synoptic page allowed them to access critical information with fewer visual transitions, reducing the need for extensive scanning across multiple displays. Some pilots mentioned that the traditional hydraulic page required more cognitive effort to interpret due to its text-heavy format, whereas the synoptic design provided a clearer, more intuitive visualization of the hydraulic system's condition. The improved efficiency in information retrieval and decision-making suggests that the synoptic page could enhance response times in real-world emergency scenarios.

Key Takeaways

1. **Enhanced Emergency Response and Performance:** The introduction of the Hydraulic Synoptic Page resulted in a more streamlined and efficient approach to handling hydraulic failures. The design supports pilots in diagnosing issues more effectively, contributing to faster and more accurate responses.
2. **Optimised Situational Awareness:** The synoptic display enhanced pilots' perception and comprehension of hydraulic system status, leading to better anticipation of system behavior and improved decision-making. The structured visualization minimized confusion and cognitive overload, helping pilots maintain high levels of awareness throughout the emergency procedure.
3. **Superior Usability and Interface Design:** The improved SUS scores indicate that the synoptic page offers a more intuitive and user-friendly experience compared to the traditional page. Reduced information clutter and enhanced graphical representation made it easier for pilots to process and react to critical alerts. These findings underscore the importance of human-centered design in cockpit interfaces, ensuring that pilots can operate efficiently under stressful conditions.
4. **Operational and Safety Implications:** Given the enhanced performance and usability, adopting the Hydraulic Synoptic Page in operational aircraft could contribute to reducing pilot workload and improving emergency management. The study's findings provide valuable insights for future cockpit interface development, reinforcing the necessity of intuitive visual displays in high-risk aviation environments.

References

Endsley, M. R., & Robertson, M. M. (2000). *Situational awareness in aircraft maintenance teams.* International Journal of Industrial Ergonomics 26 (2000) 301-325.

Hicks, J., Durbin, D., Morris, A., & Davis, B. (2014). *A Summary of Crew Workload and Situational Awareness Ratings for U.S. Army Aviation Aircraft.*

Laubheimer, P. (2018). *Beyond the NPS: Measuring Perceived Usability with the SUS, NASA TLX, and Single Ease Question After Tasks and Usability Tests.* Nielsen Norman Group.

Tayor, R. M. (1990). *Situational Awareness Rating Technique (SART): The development of a tool for aircrew systems design. In Situational Awareness in Aerospace Operations. (p. 1-17).*

Maintenance errors in commercial aviation: contextualising an undefined systemic problem

Kevin Hayes[1], Laetitia Marquie[2] & Lee Dann[1]

[1]Airbus, UK, [2]Airbus, France

SUMMARY

In the commercial aviation sector, Maintenance Errors have resulted in several high-profile accidents and incidents and the annual cost to the industry is estimated to be some US$616M (Allianz, 2024). Despite an industry wide consensus from both regulators and operators alike that the problem of Maintenance Errors needs to be addressed, there remains no definition of what a Maintenance Error is. Using both literature reviews and data from interviews with airline safety professionals, this paper explores why there has been an issue in aviation with regards to understanding Maintenance Error and proposes the Aviation Maintenance System (AMS) as a model for contextualising the problem, as well as proposing a working definition for Aviation Maintenance Errors (AMEs).

KEYWORDS

Aviation Maintenance Error, Aviation Safety, Maintenance

Introduction

In the aviation sector there has been a growing awareness over the past few decades that Maintenance Errors have a significant impact on the safety (BASI, 1997), efficiency and profitability of airlines (RAeS, 2022). Despite a shared consensus from both the regulators and the industry that this issue should be addressed, there is still no industry wide definition of what a Maintenance Error is and consequentially there is no unified approach to address the problem. This paper explores why there has been an issue in aviation with regards to understanding Maintenance Error and proposes the Aviation Maintenance System (AMS) as a model for contextualising the problem before proposing a working definition for Aviation Maintenance Errors (AMEs).

Background

High profile events such as the Air China flight 120 incident of 2007 in Naha, Japan, show how critical maintenance can be to the safe operation of commercial aircraft. This event saw the total destruction by fire of a Boeing 737-800 after a single washer was not fitted during maintenance to the downstop of a wing slat (Japan Transport Safety Board, 2009). For reasons partially explored by this paper, getting accurate global safety data regarding Maintenance Error in aviation is difficult, so the scale of the problem can be difficult to quantify and contextualise. The UK Air Accident Investigation Branch (UK AAIB) conducted an analysis of the occurrences it investigated over a 12 year period and found that 55 events, some 2% of the total, involved some sort of maintenance error (RAeS, 2022). The UK AAIB only investigates those occurrences above a certain threshold but an Airbus internal study that looked at occurrences with a much wider range of severities placed the estimate in a similar order of magnitude, with some 5% of occurrences reported to the company in 2024 containing details that could be attributed to Maintenance Error. When compared to the number of safety related events involving flight operations, the number of Maintenance Errors that directly led to an accident might seem small but these figures do not account for safety related events in which a maintenance error was not the primary cause of the accident but still may have been a significant causal factor. An example of such an accident is the PenAir flight 3296 incident in 2019, when a Saab 200 overran the end of the runway at Unalaska Airport, Alaska resulting in the death of a passenger. In this accident the brake system was found to be cross wired and while this did not lead directly to the accident it was a significant contributing factor as to why the aircraft came off the runway (National Transportation Safety Board, 2021).

Maintenance Errors also come with a significant financial cost. Not only do errors in maintenance require the task to be redone, doubling the cost of the task in terms of wages and materials, they often have knock on effects to the airlines operational schedule which can result in expensive compensation pay-outs to customers who are impacted by delays, in-flight turn backs or flight rerouting. Over a five-year period “faulty workmanship” accounted for 22% of aviation insurance claims at a cost of more than US$3.08bn (Allianz, 2024). The cost of reworking maintenance errors is set to rise as the industry faces a number of challenges such a global shortage of mechanics, increasing costs for parts and next generation aircraft/engines being more expensive to service and repair (Allianz, 2024). In an industry with notoriously tight margins the ability to control such costs is the difference between success and failure for an airline.

Doing nothing to reduce the number of Maintenance Errors in aviation is no longer an option. The worldwide aviation fleet is set to double in the next 20 years (Airbus, 2024) and even if the rate of Maintenance Error remains the same the industry will still see in real terms a doubling of the number of safety related incidents attributable to maintenance, along with a doubling of the associated cost. So, with a strong rationale as to why the phenomenon of Maintenance Error needs to be addressed the question remains, what exactly is a Maintenance Error?

Maintenance Error - the cross-sector view

The concept of Maintenance Error, as its name implies, exists at the intersection of two disciplines, that of engineering and that of Human Factors. When discussed by maintainability and reliability engineers, Maintenance Errors are usually dismissed as being the result of the wrong preventative or repair actions (Dhillon, 2006). This is because maintenance is seen by many engineers as merely being a set of actions required to maintain an item in a serviceable condition with the description of what maintenance people do often being mistaken for the intent of what those same actions are trying to achieve (Kinnison & Siddiqui, 2013).

The discipline of Human Factors takes a more systemic view of Maintenance Error and it does so in order to try and reduce the error rate by identifying the underlying causes. One of the largest explorations of the topic has been done by Reason and Hobbs (Reason & Hobbs, 2003) who discuss the various types of Human Error and how they may apply in a maintenance setting regardless of which sector the maintenance is occurring in. Despite considering many micro and macro factors that might influence Maintenance Error, at no point is a definition of a Maintenance Error given. Across the body of Human Factors and Ergonomics literature this tendency is repeated, a keyword search of "Maintenance Error” in four leading HFE Journals found a number of papers that discuss how Maintenance Errors can be investigated or reduced. Those few papers that used the term "Maintenance Error” do not provide a definition of what this phrase means but would discuss it in terms of Human Error in a maintenance environment and usually with reference to Rasmussen's Skill-rule-knowledge model (Rasmussen, 1983) and Reasons additional work on this framework describing slips, lapses etc (Reason, 1991).

Maintenance Error in Aviation

The regulatory perspective. The two largest and most influential regulators in commercial aviation around the world are the Federal Aviation Administration (FAA) in the USA and the European Union Aviation Safety Agency (EASA). Despite acknowledging the safety concerns that arise from Maintenance Errors (EASA, 2024) and repeatedly referring to Maintenance Error in the regulatory design requirements (EASA, 2023) neither organisation provides a definition of what a Maintenance Error is. How has such a situation arisen and why is it important?

EASA has very detailed and specific regulation regarding the impact of cockpit design on the performance of pilots (EASA, 2023). This regulation and the subsequent work of HF professionals to try and reduce pilot error as much as possible has widely been credited with making modern aviation as safe and reliable as it is today (Mathavara et al, 2022). While EASA, as part of its plan for aviation safety, is currently trying to emulate the success of HF in the cockpit by running an initiative to better explore how to mitigate against Maintenance Errors through design (EASA, 2024) there remains no single regulation that requires HF to be considered in the design of maintenance tasks in the same way that HF is considered in the operation of an aircraft (EASA, 2023). Instead within CS25, the rules for designing large aircraft, there are multiple references to maintenance errors and the need for the design of certain systems to be "tolerant" of them, for example the wiring systems around fuel tanks (EASA, 2023). This can be attributed to the regulation evolving over time and in response to specific safety related occurrences. While there is nothing intrinsically wrong with this approach it does result in the non-consistent application of HF best practice across the aviation maintenance domain. What's more, without an agreed upon working definition of Maintenance Error how can designs be made "tolerant"? A problem cannot be solved without knowing what the problem is, and as such the first step in tackling Maintenance Errors should be to define what they are.

The perspective of airline operators. Away from the regulatory sphere, there have been a few attempts in the past at defining what a Maintenance Error means in an aviation context, though there remains no industry wide consensus with no definition being adopted nor in common use. One example that is fairly typical comes from Marx and Graeber (Marx, 1997) who define a Maintenance Error in an aviation context as "*an unexpected aircraft discrepancy attributable to the actions of an Aviation Maintenance Technician*". This definition is problematic insofar as it places the responsibility for committing the error firmly on the actions of the individual and so ignores the systemic nature of Maintenance Error as identified by Reason and Hobbs (Reason & Hobbs, 2003). This emphasis on the actions of the Maintenance Technician is a recurring element in commercial aviation and one of reasons why the issue of Maintenance Error has not been addressed in a systemic way.

The Boeing Maintenance Error Decision Aid (MEDA) model is the most ubiquitous model in use in aviation that is used solely for the analysis of Maintenance Error. While MEDA acknowledges the systemic nature of Maintenance Errors (Boeing, 1995) its classification system lends itself towards misunderstanding by users and toward an emphasis on the actions of the Maintenance Technician. By way of example, MEDA defines Maintenance Error as "*the error that directly leads to the event*" and specifies seven different error types including, Installation Error, Servicing Error and Repair Error. All of these error types create a framing that creates a causal link between the actions of the individual and the unintended event/occurrence and so detracts the organisation from looking for systemic solutions to their recurrent issues.

To better understand the issues faced by airlines when conducting analysis of in-service events that involve maintenance error, Airbus engaged with stakeholders and safety investigation professionals from six different major airlines that had a combined fleet of over 800 passenger planes. As part of this research 24 semi-structured interviews were conducted by videoconference and lasted between 1.30hr and 2hrs each with the results being analysed using Thematic Analysis (Braun & Clarke, 2006). It was found that airlines often struggled to mitigate against maintenance error recurring. Tools such as MEDA and HFACS would cause them to confound the symptoms of what went wrong (i.e. that the item was incorrectly repaired) with what the actual failing was within the socio-technical system that allowed such a symptom to manifest (i.e. why the item was incorrectly repaired). These findings would then be represented in the metrics of the airlines since the trend analysis of error was found to be closely linked to the classification methodology being used. By way of example, every airline reported that the most common maintenance error that they experienced was some variation on "component not installed correctly". This finding is a common one within the literature as well, with similar observations being made by Rankin & Sogg (2003) who were also analysing MEDA data. Knowing that the most common maintenance error in their organisation was parts being misinstalled was found to not be very useful to the Airlines as there are many Performance Shaping Factors (Kirwan, 1994) that could lead to a misinstallation and as such it was extremely difficult to implement effective mitigations to avoid repeat occurrences.

It should be reiterated that the MEDA methodology does take a systemic approach to addressing maintenance errors, unfortunately the evidence shows that many organisations across aviation struggle to apply MEDA in a way that allows them to effectively mitigate against repeat maintenance errors. Part of this problem is that these organisations do not understand the systemic nature of maintenance errors and/or do not have analytical tools that match their needs.

The systemic context

This paper has referred several times to the systemic nature of Maintenance Error, but what is this system that is being alluded to? Before attempting to provide a definition for Maintenance Errors in the aviation domain we must first consider the socio-technical system in which that maintenance is being conducted. This will help with contextualising the problem and help with understanding why it is valid to refer to Aviation Maintenance Errors (AME's) as their own phenomenon, with their own definition, rather than referring to Maintenance Errors that just so happen to be occurring in an Aviation environment.

If an attempt was made to describe in detail the systems-of-systems that covers the topic of maintenance, regardless of which domain that activity is occurring in, it is clear that the size of the system being described would be vast. Such a description would have to cover all maintenance activities on all human inventions that currently exist and would include everything from the domestic goods in a home to the International Space Station. The description of such an open system would either be so large as to be impractical, or so generic as to not give any useful insights. In contrast, aviation maintenance is a highly regulated field and as such it is a much more closed system that can be better defined and bounded. Such a well-defined boundary allows for the possible identification of common Performance Shaping Factors within that system that could contribute to Maintenance Errors occurring. Additionally, these Performance Shaping Factors could be of sufficient specificity and granularity to facilitate organisations with the identification of trends and the development of robust mitigations. In the aviation maintenance environment, a model depicting such a system-of-systems can be usefully thought of as an Aviation Maintenance System (AMS).

Figure 1 illustrates the AMS that governs how aviation maintenance is conducted in an organisation bound by EASA regulations. It shows that an AMS truly is a system-of-systems as it meets all five of the basic criteria proposed by Maier (1998); independence, evolutionary development, possessing emergent behaviour and being geographically distributed. This model shows that despite the complexity of the AMS, the system is very well defined with regulations setting standards in all three categories of Performance Shaping Factor as suggested by Galyean (2006); the individual, the organization, and the environment. For the individual PSF's we can see that not only are there standards for the training of mechanics and licenced engineers (Part 147 and Part 66) but also responsibilities laid upon the Part 145 organisation around how these standards are monitored and assured. For the organisational PSFs there are standards for the production of procedures (Part 145 and Part M) and how effectively these procedures are applied (Part M). For environmental PSFs, maintenance personnel can expect themselves to be working in a hangar of sufficient quality (Part 145) or on an aircraft built to a known standard (Part 21) and using standardised tools and spares (Part 145 and Part 21). While this AMS only covers maintenance conducted in the regulatory sphere of EASA, similar AMS's could be produced for organisations in other regulatory areas such as those covered by the FAA. Due to international agreements between regulators, such as the bilateral Aviation Safety Agreement between EASA and the FAA, it can reasonably be expected that models of other AMS's will just as comprehensively cover the Performance Shaping Factors that influence a maintainer conducting aviation maintenance.

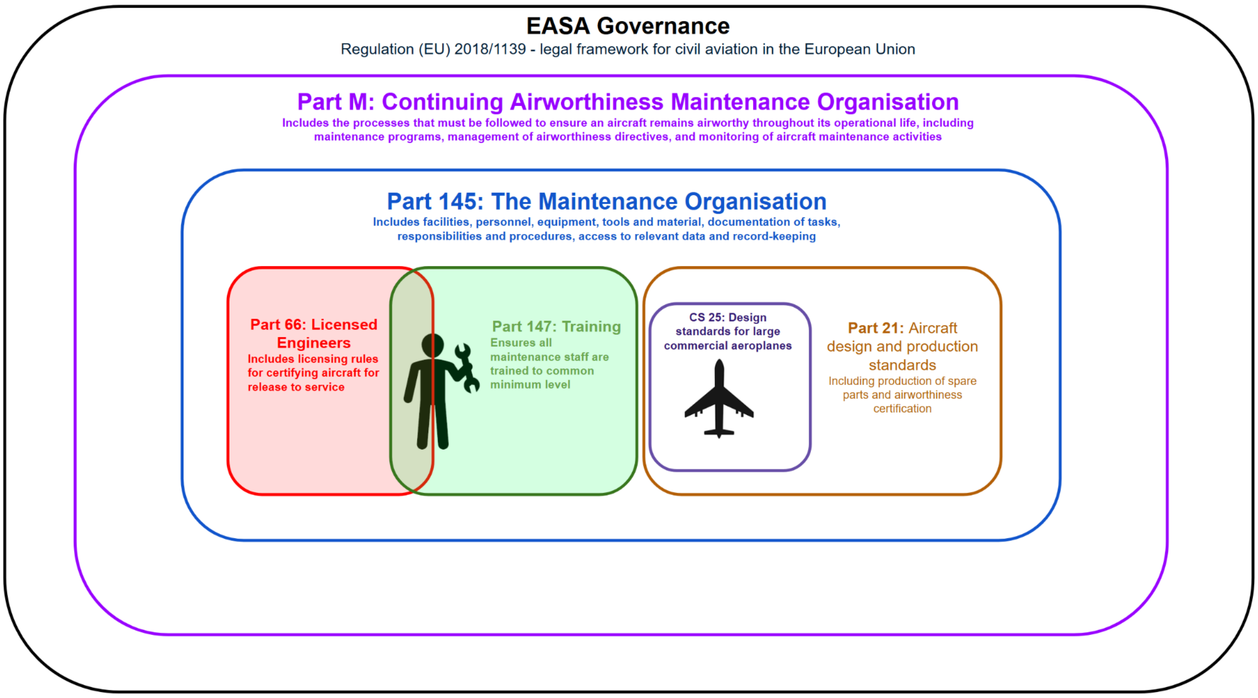


Figure 1: The Aviation Maintenance System (AMS) for an EASA approved maintenance organisation

Aviation Maintenance Error

Through the model of an AMS it now becomes easier to define Maintenance Error, but before doing so we must first consider what is maintenance and what is its purpose? *Maintenance* is defined as the *process* of ensuring that a system continually performs its intended function and at its designed-in level of reliability and safety (Kinnison & Siddiqui, 2013). Defining maintenance as a process shows that there is a difference between the word when meant in an engineering or technical sense versus when the same word is colloquially or commonly used to describe the act of what a maintainer does when carrying out their job role. An AMS *is the process* by which an aircraft is ensured to perform with both high safety and reliability. The AMS covers all aspects from when the aircraft was designed through to how it is maintained, who does the maintenance and where that maintenance is conducted. It includes all interdependencies and complexities that are required as part of that process occurring within a socio-technical system. It therefore becomes clear that when a Maintenance Error occurs in an aviation environment that something beyond the actions of the individual must also have contributed to the error occurring since the goal of the entire AMS is to avoid such a degradation of performance. It therefore follows that if it is useful to discuss the system which governs aviation maintenance in terms of an AMS, it becomes equally useful to discuss failures within that system-of-systems as Aviation Maintenance Errors (AMEs). This also helps to distinguish the concept of AMEs from the concept of maintainer error or from the term Maintenance Error when applied to non-aviation domains, which tend to be more open systems.

Based upon this reasoning, the following definition of an Aviation Maintenance Error (AME) within the context of the commercial aviation sector is therefore proposed: *When the Aviation Maintenance System (which includes people) allows, or could have allowed, a maintenance activity to be completed in a way that* ***does not*** *fully meet the* ***intended outcome*** *of the task as per the design intent or the regulatory requirements.*

Application of the definition

As previously described, many airlines struggle to mitigate against repeat occurrences of AMEs because the *symptom* of the error is often considered to be the *actual* error. By looking at a simple case study it is possible to see why this misperception is an issue and how the model of an AMS can be used as a tool to develop better solutions.

Use case. The misinstallation of components is continuously identified as the largest category of Maintenance Error identified by analysis (RAeS, 2022 and Rankin & Sogg, 2003). There are many different ways in which a misinstallation can occur, but one of the most common is through a configuration control issue. Configuration control refers to the process of ensuring that the aircraft always operates with the intended and approved configuration of components installed, so as to avoid unexpected system behaviors which could affect the safety of the aircraft. With an aircraft being made up of thousands of different parts this can be a complex task as different modifications are introduced over time and different combinations of components becomes possible.

On one occasion, an Airbus customer identified that one of their aircraft had been operating for some time with a component fitted that was not compatible with the current aircraft configuration as mandated by the approved Aircraft Technical Configuration document. The customer made a request to Airbus asking whether there was a safety concern with having this component fitted and, if not, would it be possible to add the item to the approved Aircraft Technical Configuration so as to avoid the nugatory replacement of the item. In this instance there was no safety concern and Airbus was able to approve the customer's request that the item remain fitted for the current configuration of the aircraft.

At the time, this sequence of events was not considered by the customer to constitute a Maintenance Error. As far as the customer was concerned the component was not misinstalled as there was no operational impact to the operator and consequently no further investigative action was taken. What this approach did not consider was that, despite the final outcome, there was still a latent error (Ramanujam & Goodman, 2003) within the systems-of-systems that allowed the aircraft to fly in a configuration that, at the time, was unapproved. By only focusing on the outcome, rather than on the error mechanism, there was no acknowledgement that it was only by pure luck that the aircraft could be approved to remain in this configuration. This also meant that without further action there remained the possibility that another incompatible component could be fitted to an aircraft, and that the checks and balances within the organisation may not be robust enough to immediately identify such an issue, as had previously happened. Had the operator understood the concept of AME in the context of an AMS then they would have been more likely to try and identify the contributing factors that led to this occurrence so that appropriate mitigating actions could have been put in place to prevent a similar event happening again in the future.

This use case therefore shows how the lack of an agreed definition for Maintenance Error within the aviation domain perpetrates misconceptions and contributes towards the issue not being systemically addressed.

Conclusion

This paper has attempted to provide a definition of Aviation Maintenance Errors that places the phenomenon within the systemic context of an Aviation Maintenance System. For regulators, airlines and manufacturers to effectively address AME such a definition, or something similar, must be adopted across the industry so that the issue can be properly discussed, analysed and mitigated against. Future work that would capitalise on this systemic approach would be the development of a taxonomy that allows for the identification and trending of Performance Shaping Factors at a sufficient level of granularity to allow organisations to identify the best way to mitigate against repeat occurrences of Aviation Maintenance Errors.

References

Allianz. (2024). Aviation Risk Report 2024. Available for download at: https://commercial.allianz.com/news-and-insights/news/aviation-trends-2024.html

Airbus. (2024). Global Market Forecast 2024. Available for download at: https://www.airbus.com/sites/g/files/jlcbta136/files/2024-07/GMF%202024-2043%20Presentation_4DTS.pdf

Boeing. (1995). Maintenance Error Decision Aid (MEDA) Users Guide. Available at: https://omnisms.aero/wp-content/uploads/2016/12/Boeing-MEDA-Users-Guide.pdf

Bureau of Air Safety investigation (BASI). (1997). Human factors in airline maintenance: A study of incident reports. Available at: https://www.atsb.gov.au/sites/default/files/media/30068/sir199706_001.pdf

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. Available at: https://doi.org/10.1191/1478088706qp063oa

Dhillon, B.S (2006). Maintainability, Maintenance, and Reliability for Engineers, 1st edition, Boca Raton, CRC Press. Available at: https://doi.org/10.1201/9781420006780

European Union Aviation Safety Agency (EASA). (2023).Certification Specifications and Acceptable Means of Compliance for Large Aeroplanes (CS-25) Amendment 28. Available at: https://www.easa.europa.eu/en/document-library/certification-specifications/cs-25-amendment-28

European Union Aviation Safety Agency (EASA). (2024). European Plan for Aviation Safety (EPAS) Volume III — 2024 edition, Inadequate aircraft design resulting in maintenance errors (SI-9006) page 65. Available at: https://www.easa.europa.eu/sites/default/files/dfu/annex_to_easa_mb_decision_09-2023_on_epas_-_vol_iii_2024.pdf

Galyean, W. J. (2006). Orthogonal PSF taxonomy for human reliability analysis. In Proceedings of the 8th International Conference on Probabilistic Safety Assessment and Management," Paper PSAM-0281 (pp. 1-5). Available at: https://doi.org/10.1115/1.802442.paper207

Japan Transport Safety Board. (2009). Aircraft Accident Investigation Report, China Airlines (Taiwan) Boeing 737-800, B18616, Spot 41 at Naha Airport, August 20, 2007, at about 10:33 JST. Available at:https://www.faa.gov/sites/faa.gov/files/2022-11/Okinawa_accident_report.pdf

Kirwan. (1994). A Guide To Practical Human Reliability Assessment (1st ed.). Routledge.

Kinnison, H. A., & Siddiqui, T. (2013). Aviation maintenance management (2nd ed). McGraw-Hill. https://accessengineeringlibrary.com/browse/aviation-maintenance-management-second-edition

Maier, M.W., (1998). Architecting principles for system of systems. Systems Engineering, 1 267-284. Available at: https://doi.org/10.1002/(SICI)1520-6858(1998)1:4%3C267::AID-SYS3%3E3.0.CO;2-D

Mathavara, Kamaleshaiah & Ramachandran, Guruprasad. (2022). Role of Human Factors in Preventing Aviation Accidents: An Insight. Available at: DOI:10.5772/intechopen.106899.

Marx, D.A, & Graeber, C. (1997). Aviation Psychology in Practice, Chapter 5, Human Error in Aircraft Maintenance, 1st edition 1997, Routledge. Available at: https://doi.org/10.4324/9781351218825

National Transportation Safety Board. (2019). Runway Overrun during landing. Peninsula Aviation Services Inc. d.b.a. PenAir flight 3296 Saab 2000, N686PA. Available at: https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR2105.pdf

Rankin, W. L & Sogg, S. L. (2003). Update on the Maintenance Error Decision Aid (MEDA) Process. Paper presented at the MEDA/MEMS Workshop and Seminar. May 21-23, 2003, Aviation House, Gatwick, UK.

Ramanujam, R., & Goodman, P. S. (2003). Latent errors and adverse organizational consequences: A conceptualization. *Journal of Organizational Behavior, 24*(7, Special Issue), 815–836. Available at: https://doi.org/10.1002/job.218

Rasmussen, J. (1983) Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance, IEEE Transactions on Systems, Man, and Cybernetics Volume: SMC-13, Issue: 3, May-June 1983, pages 257-266. Available at: https://doi.org/10.1109/TSMC.1983.6313160

Reason, J. (1991). Human Error, University of Manchester. Available at: https://doi.org/10.1017/CBO9781139062367

Reason, J, & Hobbs, A. (2003). Managing Maintenance Error A practical Guide, 1st edition, London, CRC Press. Available at: https://doi.org/10.1201/9781315249926

Royal Aeronautical Society (RAeS) Human Factors: Engineering Sub-Group Project. (2022). Development of a strategy to enhance human-centred design for maintenance. Available at: https://www/aerosociety.com/media/19526/human-factors-engineering-sub-group-report-oct2022.pdf

Resident Pathogens in Systems Engineering: Boeing 737 Max 8 Crashes Case Study

Sanjeev Kumar Appicharla

SUMMARY

The aim of the paper is to present "resident pathogens" or "latent failure conditions" identified in Systems Engineering (SE) practices that led to the two fatal accidents of Lion Air Flight 610 and Ethiopian Airlines Flight 302 involving Boeing 737 Max- 8 airplanes. The accidents led to the tragic loss of 346 lives and incredible pain for the victims' families.

The desktop study included collecting information and data that is publicly available to represent all relevant viewpoints to ensure completeness. The accident analysis uses the Cybernetic Risk Management Model (CRMM) with support from the hybrid Swiss Cheese Model (SCM) and Management Oversight & Risk Tree (MORT) fault tree analytical technique. Resident pathogen metaphor in the Swiss Cheese accident causation model denotes fallible decisions made during the (SE) processes. The methodology incorporates Jens Rasmussen's risk management framework (RMF) augmented by the Heuristics & Biases (H&B) approach to decision making, Thus, the methodology can help to identify latent failures conditions at all levels of the socio-technical system involved in the control of the system -of- interest (SOI).

KEYWORDS

Aerodynamics Stability Complex systems, Accident Analysis, Human, organisational and technical (HOT) Factors, Safety Risk Management

Introduction

Information about the tragic accidents is presented. This information describes the loss events- T and the accident(s) SA1 as per the MORT Manual (Kingston et al, 2009a).

The National Transportation Safety Board Report states thus: "On October 29, 2018, PT Lion Mentari Airlines (Lion Air) flight 610, a Boeing 737 MAX 8, PK-LQP, crashed in the Java Sea shortly after takeoff from Soekarno-Hatta International Airport, Jakarta, Indonesia. The flight was a scheduled domestic flight from Jakarta to Depati Amir Airport, Pangkal Pinang City, Bangka Belitung Islands Province, Indonesia. All 189 passengers and crew on board died, and the airplane was destroyed." National Transportation Safety Board Report (NTSB, 2019a).

"On March 10, 2019, Ethiopian Airlines flight 302, a Boeing 737 MAX 8, Ethiopian registration ET-AVJ, crashed near Ejere, Ethiopia, shortly after takeoff from Addis Ababa Bole International Airport, Ethiopia. The flight was a scheduled international passenger flight from Addis Ababa to Jomo Kenyatta International Airport, Nairobi, Kenya. All 157 passengers and crew on board died, and the airplane was destroyed." (NTSB, 2019a).

(KNKT, 2019) listed nine contributing factors (clause 3.2). that were related to lifecycle Systems Engineering processes of concept definition, system definition, system realisation and systems deployment (INCOSE, 2023, see Figure 2.10). (EAAIB, 2022) stated repetitive and uncommanded airplane-nose-down inputs from the MCAS due to erroneous AOA input, and its unrecoverable activation system which made the airplane dive with the rate of -33,000 ft/min close to the ground was the most probable cause of the accident (clause 3.2). The EAAIB Report listed ten contributing factors (clause 3.2). (BEA, 2023) and (NTSB, 2023) dispute the probable cause of the accident adding that inadequate pilot crew handling of the aircraft before the uncommanded Maneuvering Characteristics Augmentation System (MCAS) activation and the airport management of foreign objects (like birds striking) to be included as well in addition to MCAS related factors.

In terms of aviation occurrences classification by (ASA JSAT, 2014). these accidents are classified as Loss of Control in Flight (LOC-I) accidents. (Bromfield & Jamieson, 2022) did not note the 2014 JSAT paper.

Cognitive biases are mental errors in judgment under uncertainty caused by our simplified information processing strategies (sometimes called heuristics) and are consistent and predictable (Tversky & Kahneman, 1974), (INCOSE, 2023). Since Tversky and Kahneman's 1974 seminal paper, behavioural decision researchers have identified a large number of biases in human judgment and decision making, each showing a deviation from a normative rule of probability or utility theory (Montibeller &Von Winterfeldt, 2015). When evaluating the results of risk assessments under the managerial review and judgement process, the potential for cognitive biases and heuristics that could influence how decision-makers interpret and deliberate over the results of the risk assessments is to be addressed (Glette-Iversen et al., 2023). Cognitive biases can contribute to incidents, failures, or disasters as a result of distorted decision making and can lead to undesirable outcomes is noted by (INCOSE, 2023). Omission bias in organisations does arise when the phenomenon of emergence is neglected (Reiman & Rollenhagen, 2011). We overestimate the likelihood of good outcomes and underestimate the likelihood of bad outcomes under the influence of optimism bias (Kahneman, 2012), (Montibeller &Von Winterfeldt, 2015). The confirmation bias, or the tendency to look for (and find) information that confirms expectations and disregards information that negates them, might distort probability estimates. Considering the importance of eliciting judgments (probabilities, values, utilities, weights, etc.) in decision and risk analysis, it is somewhat surprising that relative little attention has been previously paid to the possible distortions of an analysis due to these biases(Montibeller &Von Winterfeldt, 2015).

The importance of systemic and organisational performance shaping conditions has been clearly established in the safety literature (Starbuck & Farjoun (Eds.), 2005), (van Kampen, J., et al., 2017), (INCOSE HSI Working Group, 2023). But it is challenging for practitioners from Systems Engineering (including human factors, safety specialists and traditional systems engineers) discipline to identify latent failure types and to find ways to neutralizing the pathogens revealed to improve safety performance (Reason, 1990b) (pp.210-11), (Appicharla, 2006a), (Macrae, 2007). Social and institutional factors are involved in safety risk management (Hutter & Jones, 2006), (Reason, 1990b) (pp. 216) and their contribution to safety risk management is noted (Macrae, 2007). Reason (1990, 1997) has argued that unknown, latent risks are inevitable in all organizations and the primary purpose of risk analysis is to find them and "make them visible" (Reason, 1997, p. 37), (Macrae, 2007). Section 2.3 of the ICAO website "Safety Management System Implementation" states on Accident Causation, thus: "Safety risks can be generated by active failures and latent conditions. The concept of accident causation is an active field of study, and many types of models exist to illustrate the events taking place leading up to an accident." (ICAO, 2018). For definition of active failures and latent failure conditions, (Appicharla, 2023a) may be consulted.

The aim of this paper is to highlight resident pathogens such that systems engineers may identify, reflect and improve the design and delivery process. A significant contribution of the paper is to highlight contribution of cognitive biases underlying decision failures and their contribution to the crashes.

The Methodology for Accident Analysis: The Cybernetic Accident Risk Management Model

At the time of submission for review, Google scholar's search yielded 15,500 hits on the theme of Boeing 737 Max 8 crashes. Due to space constraints, observations on these cannot be presented. However, it is to be noted that the systems approach is underpinned by the idea that safety is an emergent property of socio-technical systems. Interactions within the system are non-linear and produce emergent behaviors that are hard to predict (Rasmussen,1997), Rasmussen' 1997 risk management framework describes various systems levels including: government; regulators; company management; staff; and work. According to Rasmussen, each level is involved in the management and performance of the system. The system requires "vertical integration" to maintain control of hazardous processes and create safe performance.That is, decisions made at the higher levels need to be filtered down to the lower levels and influence practice, and equally feedback from the lower levels needs to filter up and inform the decisions and actions occurring within the higher levels of the system (Grant et al, 2015). These levels need to include professional engineering associations, industry bodies and consultanices as well. (Ball & Boehmer-Christiansen, 2002) noted that many institutions and their associated professions have carved out their own specific approach to safety decision making, sometimes in isolation from other professions, and these are in many cases not consistent with each other.

Systemic view of accident causation is the norm where agencies such as the Australian Bureau of Air Safety Investigation (BASI) and the Transportation Safety Board of Canada have seen the great opportunity afforded by James Reason's organisational accident approach to identify deeper systemic causes (Braithwaite, 2010). The BASI was the first to use Reason's model for all its major reports, directing attention to organisational factors underlying aviation accidents (Hudson, 2003). The idea that human error is a symptom of system failure may be traced back to Justice Peter Mahon's examination of the circumstances behind the loss of an Air New Zealand DC10 aircraft with 257 fatalities on the slopes of Mount Erebus in 1979 (Braithwaite, 2010). Dr Rob Lee, Director of BASI, Captain Dan Maurino, Jean Paries, Captain Jeremy Butler, and Captain Bertrand de Courville are few professionals who are credited with promoting the Swiss Cheese Model in the aviation sector (Reason,J., et al, 2006), (Appicharla, 2006a). Reason argued that the safety level of any organisation could be evaluated from a limited set of indicators, based on an analogy between the breakdown of complex technological systems and the aetiology of multiple -cause illnesses such as cancer and cardio-vascular diseases(Reason, 1990b), (Larouzee & Le Coze, 2020). Like cancer or heart disease, industrial accidents result from a combination of factors, each of which is necessary but not sufficient to overcome the technical, human and organisational defences of the industrial system defences (Reason, 1990a). Criticism of the Swiss Cheese Model were taken up by (Reason et al, 2006) and answered. One of the criticisms is that misapplication of the model can shift the blame backwards, from a 'blame the pilot' to 'blame the management' culture and may obscure real human factors concerns at the front line. However, as (Reason et al, 2006) argued that the fact that deterministic causal connection between latent conditions and accidents cannot easily be identified (particularly before the event), does not rule out that efficient prevention policy can be based on addressing latent conditions. Drawing upon convergence between disciplines of decision research, organisational theory, safety management and accident causation identified by (Rasmussen,1997) and assuming the hypothesis that system safety, accident analysis and occupation safety research need a common approach (Appicharla, 2006a) used the MORT to identify latent failure conditions to meet the challenges of the Railway Safety Directive 2004/49/EC (No longer in force) (EU, 2004) in response to an internal research brief by RSSB for a Systems Engineering solution to identify missing safety measures at the duty-holders interfaces. (Appicharla, 2010c) revealed the results of application of the Methodology over five-year period at RSSB in the form of the latent failure conditions that can contribute to future railway accidents like the lockout protection system, ABCL types of level crossing, permissive working operations, axle counters and train detection systems. These hazard warnings were neglected and safety critical incidents as identified in the Hazard Analysis Reports manifested in the railway operations later. One of them accident(s) foreseen was analysed by (Appicharla, 2011) and the 2011 paper presented failings at the regulator level as well. Later, in 2017, the Heuristics and Biases (H&B) approach was integrated into the Methodology and (Appicharla,v2021c) presented the application of the Methodology to the 2017 Cambrian ERTMS Incident. The Regulatory and System owner awareness of Common Safety Method: Risk Assessment LTA is noted in the section 2c of the paper and this state of affairs is prevailing since 2007. (EU, 2016) (Recast of Directive 2004/49/EC) is in force and calls for integration of human, organisational ad technical factors. (The UK HSE, 2003) provides guidance how to manage the impact of organisational change on their control of the hazards for chemical plants.

The 2017 Cybernetic risk management model (see Figure 1) uses the hybrid Swiss Cheese Model (SCM) (Reason et al, 2006) and (MORT) terminology under the lens of socio-technical systems & 1997 Risk Management Framework(RMF) (Rasmussen,1997), control system theory vide barrier model (Ashby, 1956) cited in (Appicharla, 2006a),and H & B approach (Reason, 1990b) and (Kahneman, 2012) to explain accident causation. The Systems Engineering framework is made up of SCM as an explanatory mechanism (see accident equation (1)) and modelling the latent causal factors with the help of the MORT, the "resident pathogens" (including human, organisational, social, institutional and technical factors) are studied in the paper using the hybrid MORT & SCM model. The concept of emergent properties plays a significant role in accident analysis (Appicharla, 2006a), (Appicharla, 2010c). The Energy Barrier Trace Analysis (Table I) of the MORT Manual and its similarity with the Swiss Cheese Model (Reason et al, 2006) can be seen in (Appicharla, 2011) with discussion of biases like optimism bias play in decision making and neglect of other biases was noted as well.

The Accident Risk Management Model describes how the unsafe outcomes occur as a result of:

- Less than adequate (LTA) development/application of standards related to systems engineering, safety engineering, human factors, and domain related standards or omission of hazard analysis at the standard preparation, hazards not controlled by standards (Rasmussen et al,1994), (Starbuck & Farjoun, (Eds.), 2005),(Appicharla, 2006a), (Appicharla, 2010c), (Koopman, et al, 2019);
- LTA responses to the disturbances due to Heuristics and Biases(Reason, 1990b), (Reiman & Rollenhagen, 2011), (Kahneman, 2012), (Appicharla, 2023a). (Tuccio, 2011) drawing upon NASA Human Factors Experts study, showed that at least 19 aviation accidents over a ten-year period from 1991 through 2000 can be attributed to heuristics and recommended that pilot training should adopt these concepts. This open access article is a helpful reading to understand heuristics and biases from accident analysis perspective and enable us to reframe our thinking about the concept of Human error.
- LTA Business policy and its implementation and its integration with risk related policies (Starbuck & Farjoun, (Eds.), 2005). Use of pre-trained models can also increase levels of statistical uncertainty and cause issues with bias management, scientific validity, and reproducibility (NIST, 2023) (pp.38),(Koopman & Widen, 2023).
- LTA Safety Risk management practices of System Definition, Hazard Identification, Risk Analysis and Implementing Risk Controls Options, and Assurance Management(Appicharla, 2006a), (FAA , 2017), (Rasmussen,1997). Attention to management of desired emergent properties is needed and efforts to prevent unwanted or undesirable ones must be taken (Siemieniuch & Sinclair, 2014).
- LTA Learning from Accident Investigation, risk assessments and other public inquiries due to biases in accident investigation(Reason, 1990b), (Leveson.2011) and/ or LTA

analyses of past accident scenarios that do not serve to describe the socio-technical context within which accidental flow of events are conditioned and ultimately take place (Rasmussen & Svedung, 2000)(pp.17). Attribution error, Hindsight bias and Outcome bias may impact our learning of right lessons from past accidents (Appicharla, 2023a);

- Finally, LTA philosophy of ALARP decision making, LTA Oversight process/ LTA business/mission analysis (INCOSE, 2023). Affect and other heuristics may have an adverse impact on ALARP decision making by injecting biases (Ale et al, 2015), (Langdalen et al, 2020). Does the risk exceed an acceptable level (e.g., regulatory standards, action levels)—the "As Low As Reasonably Practical" (ALARP) test? (Smith, 2013). (The UK HSE, 2001) recognises any informed discussion on the risk decision-making process quickly raises ethical, social, economic and scientific considerations(clause 13). (Adam & Thompson, 2002) concluded attempts to manage risk that a) ignore the rewards of risk taking, and/or b) exclude significant stakeholders, and/or c) fail to appreciate the type of risk it is sought to manage, are unlikely to succeed.

Accident Analysis Results

In accordance with (ASA JSAT, 2014). the SCM and the MORT EBTA, and using the accident analysis procedures described in the MORT User Manual, and following the application examples stated in (Appicharla, 2021c) and (Appicharla, 2024b), we describe the accident(s), thus:

> Loss of Control – Inflight (LOC-I) hazard (SB1) + loss of 346 lives (SB2) + Less than adequate control by Socio-technical system in control (Omissions & oversight)

The following are significant latent failure conditions that led to the fatal crashes:

MORT code S/M. Oversight and Omissions (Kingston et al, 2009a) (pp. xvi -1): LTA Specific Control Factors) S-1[1] & LTA Management System Factor: M-37: LTA regulatory oversight & functioning, and certification): Excessive Federal Aviation Administration (FAA) delegation of certification functions to Boeing on the 737 MAX eroded FAA's oversight effectiveness and the safety of the public. Boeing's Authorised representatives (ARs)—Boeing employees acting as representatives of the FAA or performing certification functions on behalf of the FAA—were impaired from acting independently of the company about the certification of the 737 MAX(Defazio & Larsen, 2020)(pp.57). "DOT OIG Audit Report AV-2016-001 stated "FAA Lacks an Effective Staffing Model and Risk-Based Oversight Process for Organisation Designation Authorization (ibid)(pp.59). Technical design flaws, faulty assumptions about pilot responses, and management failures by both Boeing and the FAA played instrumental and causative roles in the chain of errors that led to the crashes …that resulted in the tragic and preventable deaths of 346 people. Both crashes involved Boeing 737 MAX airplanes (ibid)(pp.5). The latent failures investigated at Boeing Board level were presented by (Appicharla, 2023a). Omission bias (Reiman & Rollenhagen, 2011), Confirmation bias due to AH and inadequate mental model of the regulatory problem space (Reason, 1990b) is asserted here.

MORT Code MA3 Risk Management System LTA/ MB3. Risk Analysis Process LTA. a1. Concepts and Requirements LTA. b5-45. Specification of Requirements LTA: c11-45(Kingston et al, 2009a) (pp.45): System (1)-engineered system (737 Max-8 aircraft), system (2)-the Boeing SE Life Cycle Project Management System, and system (3)- the Boeing Enterprise Process and Innovation System monitoring, the SE Life Cycle Project Management failed. LTA adequate system definition made up of three systems (INCOSE, 2023) (section 1.3.4). Overconfidence bias is asserted here (Reason,

[1] The numbers after the hyphen refer to the MORT Manual page numbers and letters with numbers before hyphen refer to the MORT Manual nodes/code.

1990b)(89). LTA Safety standards: (Lopes, 2024) uncovered four main limitations in safety assessment guidance that contributed to the accidents: (a) limited integration of human factors and safety, (b) limited guidance for identifying assumptions, (c) limited ability to capture non-failure based causal scenarios, and (d) limited ability to understand complex nonlinear causal relationships.

-do- SB3-5 Barriers & Controls LTA : ***SC1-5 Control of Work & Process LTA: SD1-5 Technical Information LTA:a2-8 data collection LTA: b5-8 Use of Previous Accident/Incident Information LTA? b6-8 Learning from employee/contractor's personnel experience LTA :*** The FAA TARAM analysis showed that even with the FAA's Emergency Airworthiness Directive but without a fix to MCAS, there could be more than 15 fatal 737 MAX crashes over the estimated 30-year lifetime of the fleet, then estimated to be 4,800 aircraft, resulting in over 2,900 deaths Defazio & Larsen, 2020)(pp.210). Boeing did not learn that its design was flawed, or that it had made mistakes, but blamed industry-wide assumptions regarding pilot response time (ibid) (pp.231). Subject matter experts were overruled in some cases(Cantwell, 2021). This shows 'insensitivity to predictability due to RH' (Kahneman, 2012), (Appicharla, 2023a). Even after the fatal Lion Air crash, Boeing maintained that its "rationale" for removing references to MCAS from the 737 MAX training manual was still "valid and Boeing asserted that the addition of MCAS on the 737 MAX did "not affect pilot knowledge, skills, abilities, or flight safety (ibid))(pp.27). In 9 of the 18 events, flightcrew training played a role (ASA JSAT, 2014) (pp.4). The past accidents such as AF447 (2009) (Oliver et al., 2017), and TK 1951 (2009) (Appicharla, 2023a) highlight the cases where the pilot(s) reliability in handling the stall situation would be critical. AF447 (2009) shows Inappropriate Control Inputs (Oliver et al., 2017), and TK 1951 (2009) shows effects of single sensor-based architecture (Appicharla, 2023a) leading to Inappropriate Control Inputs as a contributing factor(s) (ASA JSAT, 2014) (pp.5).

-do- a2. Design and Development LTA: b9. Human Factors (Ergonomics) Review LTA: c29-49. Did not Predict Errors: LTA consideration of Human Factors in system design, operation and maintenance, and certification processes: (INCOSE, 2023) (2.3.4.6)(3.1.4) Human Systems Integration LTA. Automation Confusion/Awareness (ASA JSAT, 2014): The FAA has provided guidance that pilots should be able to respond to uncommanded MCAS activation condition within four seconds (Defazio & Larsen, 2020) (pp.24). Unfortunately, had the EICAS been installed on the Lion Air or Ethiopian Airlines flights, some experts believe it may have helped to alleviate pilot confusion—a contributing factor in both of those accidents(ibid)(pp.47). The Boeing OMB did not indicate to flight crews that they may experience multiple alerts at once leading to cognitive confusion and mental overload, often referred to as the "startle effect" (ibid)(pp.195). LTA Flight Crew Alerting: Boeing did not make the angle of attack (AOA) Disagree alert functional to indicate the significant difference between the two AOA sensors. As a result, Boeing did not contribute to adequate situational awareness of Lion Air Flight 610 piloting crew(ASA JSAT, 2014)(pp.5). (NTSB, 2006) stated that the functional implications of failures that could result from human interaction with airplane systems and components are not analyzed in safety assessments. The Safety Board is concerned that human interaction failures are not addressed in the assessment of safety-critical system. The FAA March 2002 and the HF Guidance by (Yeh et al 2016) cited in (NTSB, 2019b) and (ASA JSAT, 2014) findings were omitted by both FAA and Boeing. (Leggett, 2020) described the four themes of regulatory failures, system design, and pilot training crisis including a doctoral thesis of Dr Karlene Petitt. She is described as an experienced pilot based in the US. She has become a vocal critic of airline safety culture, drawing on research carried out for her doctoral thesis. Omission/Out of sight out of mind bias due to AH is asserted here(Reiman & Rollenhagen, 2011), (Reason, 1990b).

-do- b8-46 Energy Control LTA: c24-47 Controls and Barriers LTA: b22-53 Design Acceptance & Change Control Process LTA: c38-53 Engineering Studies LTA: Safety Risk Management LTA: The

Single & Multiple Failure (S& MF) analysis process LTA: The MORT Fault tree does not have a tree branch to deal with the safety culture and safety management aspects under the a2 Design and development processes is to be noted. Therefore, b22 -53 and c38-53 branch are considered for Safety Risk Management aspects. The S &MF analysis process was led by a Boeing Systems Engineering team along with relevant stakeholders was carried out (NTSB, 2019a). The failure analysis was "completed prior to the design change to MCAS control law during flight test and not reevaluated." This did not entail a "process violation or non-compliance (Defazio & Larsen, 2020)(pp.209). These controllers (decision makers) who from their local perspective strove to meet their programme objectives of 'Operational commonality' and 'Amended certification (ATC) requirements prepared the latent pathway(s) to accident(s) (Defazio & Larsen, 2020), (Reason, 1990b). Confirmation bias to program objectives of 'Operational Commonality', 'Energy Efficiency' and 'Amended certification (ATC)' is asserted here, (Reason, 1990b), (Teal, 2014), (Montibeller &Von Winterfeldt, 2015).

-do- b22-53 Design Acceptance & Change Control Process LTA: c43-54 Change Review Procedure LTA: In 2011, facing a competitive threat from Airbus's new, more fuel efficient, single-aisle A320 aircraft, Boeing believed it did not have time to create a new plane from scratch (Project Yellowstone was shelved under customer pressure) (Defazio & Larsen, 2020) (pp.39). LTA business and mission analysis & management (INCOSE, 2023) (2.3.5.1). (NTSB, 2019a) documents how Aerodynamic stability and statutory requirements drove the Boeing team to re-apply the LTA MCAS solution. "Boeing failed to appropriately classify MCAS as a safety-critical system, concealed critical information about MCAS from pilots, and sought to diminish focus on MCAS as a "new function" in order to avoid increased costs, and "greater certification and training impact." (Defazio & Larsen, 2020). Boeing had tremendous financial incentives to ensure the MAX program met this goal. In December 2011, Boeing agreed to pay Southwest Airlines $1 million per MAX airplane that Boeing delivered to Southwest if its pilots were unable to operate the 737 NG and 737 MAX interchangeably due to any reason. In addition, if the FAA required more than 10 hours of pilot training and/or required flight simulator training, Boeing would reimburse SWA for any direct training expense that exceeded 10 hours Defazio & Larsen, 2020)(foot notes -814 to 821).

LTA Systems Engineering application: LTA (MCAS) Systems Knowledge; invalid source data (ASA JSAT, 2014)(pp.4): The MORT Fault tree does not have a tree branch to deal with the Systems engineering and Systems thinking aspects under the a2 Design and development processes is to be noted. Based on the admission of John Hamilton, the then-Chief Engineer, that one of the two MCAS design requirements (no objectionable interaction with the piloting of the airplane and not interfere with dive recovery) were not met, the Report concluded that MCAS was poorly designed, not adequately tested, and had received flawed oversight by the FAA (Defazio & Larsen, 2020)(pp.120). The huge error of omission is that Boeing failed to disclose the existence of MCAS to the pilot community(Defazio & Larsen, 2020) (pp. 206). Omission bias due to AH is concluded (Reason, 1990b). (Campbell, 2019) noted that Boeing 737 Max 8 is a perfect example of the cross purposes at which business, technology, and safety often find themselves. With its bottom line threatened, Boeing focused on speed instead of rigor, cost-control instead of innovation, and efficiency instead of transparency. The FAA got caught up in Boeing's rush to get the Max into production, arguably failing to enforce its own safety regulations and missing a clear opportunity to prevent these two crashes.

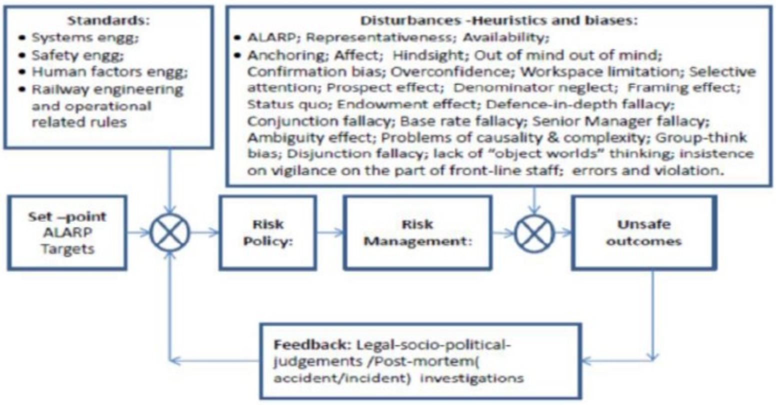


Figure 1: The Cybernetic Risk Management /Accident Risk Management Model

LTA Systems Thinking concepts and their application: Concepts of emergent properties (Hitchins, 2007), like system safety (Appicharla, 2006a), (Leveson, 2011), affordance of harm (Appicharla, 2011), inadequate communication of Systems Knowledge(ASA JSAT, 2014), and Cognitive biases (such as rankism in decision making) (Jackson, 2018) to assure better situational awareness, design and control (Defazio & Larsen, 2020), (Checkland,1981) were not part of shared mental models (Senge,1990b) of the FAA and Boeing Systems Engineering teams. Omission and confirmation biases due to AH and RH is asserted here(Reason, 1990b).

Brief discussion: The paper did not discuss the active failures of flight crew and the debate between the BEA, NTSB and the EEAIB on the question of flight crew errors due to space constraints. The role of deadheading pilot in previous flight to the Lion Air Crash is discussed by (Defazio & Larsen, 2020) and other inadequacies in maintenance activity by (KNKT), 2019). (Jackson, 1997) may be consulted to learn how the Integrated Project Teams (IPT) and other SE concepts are linked to certification and design process. (Defazio & Larsen, 2020) provided evidence for the confirmation bias hypotheis when the Boeing's own simulator tests showed that pilot response far exceeds(ten seconds in the case of one pilot) the four seconds requirements and leads to catastrophic results.

Conclusion

Resident pathogens identified in Systems Engineering (SE) practices that led to the two fatal accidents that can be addressed were presented.

Acknowledgement

The author thanks the peer reviewers for review suggestions. The author also expresses gratitude to immediate and extended family members as well. Gratitude is expressed towards safety scientists and human factors experts- Jens Rasmussen and James Reason as well.

References

Adam J & Thompson, M. (2002). UK HSE 035: Taking into account societal concerns about risk : Framing the Problem. Norwich: The UK HSE. Retrieved December 2015, from http://www.hse.gov.uk/research/rrpdf/rr035.pdf.

Ale, B. J. M. et al. (2015). ALARP and CBA all in the same game. Safety science, 76, 90-100. Retrieved April 21st, 2021, from https://www.sciencedirect.com/science/article/abs/pii/S0925753515000405

Appicharla, S. K. (2021c). Case Study Analysis of 2017 European Railway Traffic Management Incident: The Application of SIRI Methodology. World Academy of Science, Engineering and Technology International Journal of Industrial and Systems Engineering, 16(11), 15. Retrieved October 10th, 2021, from World Academy of Science, Engineering and Technology: https://publications.waset.org/10012798/case-study-analysis-of-2017-european-railway-traffic-management-incident-the-application-of-system-for-investigation-of-railway-interfaces-methodology

Appicharla, S. K. (2010c). System for Investigation of Railway Interfaces. 5th IET International Conference on System Safety (pp. 1-6). Manchester: IET. Retrieved from https://ieeexplore.ieee.org/document/5712351

Appicharla, S.K. (2024b). Accident Case Study Analysis of 2018 Developmental Automated Driving System Collison. ANNUAL INTERNATIONAL SYSTEM SAFETY SUMMIT AND TRAINING 2024 (p. 20). Minneapolis, MN 55403: INTERNATIONAL SYSTEM SAFETY Society. Retrieved Jan 9th, 2025, from https://bit.ly/3PqzTSc

Appicharla, S.K.,RSSB. (2006a, June 7th). System for Investigation of Railway Interfaces. 1st IET International Conference on System Safety: (pp. pp.7-16). London: Institution of Engineering and Technology. Retrieved December 2015, from IEEE Xplore: https://bit.ly/3Mkak1I

Appicharla,S. (2011, September 21st). Modelling and Analysis of Herefordshire Level Crossing Accident using Management Oversight and Risk Tree (MORT). The Sixth IET International System Safety Conference, (p. 10). Birmingham, UK: IET. Retrieved April 27th, 2023, from IEEE Explore: https://ieeexplore.ieee.org/document/6136924

Appicharla,S.K. (2015a, October 27th). CROSS RAIL TRAIN PROTECTION (PLAN B) - RAILWAY SAFETY REGULATIONS. Retrieved May 18th, 2019, from ORR.gov.uk: https://bit.ly/3IXXGEk

Appicharla,S.K. (2022a, April 19th). *From Nobel Prizes to Safety Risk Management: How to identify Latent Failure Conditions in Risk Management Practices.* https://www.intechopen.com/chapters/81390

Appicharla.S. (2023a, January 30th). The Boeing 737 MAX 8 Crashes, System-based Approach to Safety —A Different Perspective. Retrieved February 6th, 2023, from https://scsc.uk: https://scsc.uk/journal/index.php/scsj/article/view/18

ASA JSAT. (2014, June 17th). Airplane State Awareness (ASA) Joint Safety Analysis Team (JSAT) Report. Retrieved November 30th, 2023, from skybrary: https://skybrary.aero/sites/default/files/bookshelf/2999.pdf

Ashby W.R PhD. (1956). An introduction to cybernetics. Retrieved January 10th, 2006, from http://cleamc11.vub.ac.be/REFERPCP.html.: http://pespmc1.vub.ac.be/ASHBBOOK.html

Ball, D.J & Boehmer-Christiansen, S. (2002). UK HSE Research Report 0034: Understanding and responding to societal concerns. Retrieved March 12th, 2021, from https://www.hse.gov.uk/Research/rrhtm/rr034.htm

BEA. (2023, January 3rd). Comment on the Final Report of the Accident to the Boeing 737 ET-AVJ and operated by Ethiopian Airlines on 10 March 2023 [Investigation led by EAIB / Ethiopia]. Retrieved January 4th, 2023, from https://bit.ly/3vYutUF

Braithwaite, G.R. (2010). Twenty-First Century Reliability Accident investigation and the search for systemic failures. Safety and Reliability, 30(3), 47-57. Retrieved December 21st, 2024, from https://www.tandfonline.com/doi/abs/10.1080/09617353.2010.11690914

Bromfield, M & Jamieson. (2022). Lion Air JT610 Boeing 737 Max 8 accident – human factors analysis. Contemporary Ergonomics & Human Factors Conference 2022. (pp. 301-320). Birmingham, : Chartered Institute of Ergonomics & Human Factors, UK,. Retrieved December 3rd, 2024, from https://research.birmingham.ac.uk/files/168927614/Bromfield_Jamieson_Lion_Air_JT610_Short_Paper_v2_final_submitred_210222_preprint.pdf

Campbell.D. (2019, May 2nd). *Redline. The many human errors that brought down the Boeing 737 Max.* (Vox Media, Inc) Retrieved April 12th, 2023, from The Verge.com: http://bit.ly/3MD2RxL

Cantwell. M, Chair. (2021, December 13th). Aviation Safety Whistleblower Report. Washington DC, 20510: The Senate Committee on Commerce, Science, and Transportation. Retrieved November 29th, 2023, fromhttps://www.commerce.senate.gov/services/files/48E3E2DE-6DFC-4602-BADF-8926F551B670

Checkland. P. (1981). Systems thinking, systems practice (1988 ed.). Chichester, Sussex, UK: J. Wiley & Sons. Retrieved February 21st, 2023, from https://archive.org/details/systemsthinkings00chec/page/244/mode/2up

Dallat, C., Salmon, P. M., & Goode, N. (2019, Nov). Risky systems versus risky people: To what extent do risk assessment methods consider the systems approach to accident causation? A review of the literature. *Safety Science, 119*, 266-279. Retrieved Feb 15th, 2025, from https://www.sciencedirect.com/science/article/abs/pii/S0925753517305295

Defazio, P.A., Larsen,R. (2020, September). Final Committee Report : The Design, Development & Certification of Boeing 737 Max Report. Washington DC: The House Committee on Transportation and Infrastructure. Retrieved January 10th, 2023, from http://bit.ly/3X46ZJG

Dekker,S. (2009, July 2nd). REPORT OF THE FLIGHT CREW HUMAN FACTORS INVESTIGATION. Retrieved May 15th, 2022, from https://www.onderzoeksraad.nl: http://bit.ly/3GyrcjF

Donaldson. W. (2017, September). In Praise of the "Ologies": A Discussion of and Framework for Using Soft Skills to Sense and Influence Emergent Behaviors in Sociotechnical Systems. *Systems Engineering, 20*(5), 467-478. Retrieved April 10th, 2023, from https://doi.org/10.1002/sys.21408

EAAIB. (2022, December 23rd). INVESTIGATION REPORT ON ACCIDENT TO THE B737-MAX8 REG.ET-AVJ OPERATED BY ETHIOPIAN AIRLINES. Retrieved January 4th, 2023, from https://bea.aero/en/investigation-reports/notified-events/detail/accident-to-the-boeing-737-registered-et-avj-and-operated-by-ethiopian-airlines-on-10-03-2019-near-bishoftu-investigation-led-by-eaib-ethiopia

EU. (2004, April 29th). Railway Safety Directive 2004/49/EC(No longer in force). Official Journal of the European Union, 44-113. Retrieved April 5th, 2010, from https://eur-lex.europa.eu/oj/direct-access.html

EU. (2016, May 11th). DIRECTIVE (EU) 2016/798 on railway safety (Recast of Directive 2004/49/EC). Official Journal of the European Union, 102-149. Retrieved December 10th, 2024, from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016L0798

Evans A.E. (2020, May). *FATAL TRAIN ACCIDENTS ON BRITAIN'S MAIN LINE RAILWAYS.* Retrieved December 11th, 2023, from https://www.imperial.ac.uk/people/a.evans/document/7282/FTAB2019/?FTAB2019.pdf

FAA . (2017, May 2nd). Safety Risk Management Policy, FAA Order 8040.4B. Retrieved July 1st, 2021, from https://www.faa.gov/documentLibrary/media/Order/FAA_Order_8040.4B.pdf

Glette-Iversen, I., Flage, R., & Aven, T. (2023). Extending and improving current frameworks for risk management and decision-making: A new approach for incorporating dynamic aspects of risk and uncertainty. Safety science, 168(106317), 1-14. Retrieved Feb 5th, 2025, from https://www.sciencedirect.com/science/article/pii/S092575352300259X

Grant, E., Goode, N., & Salmon, P. M. (2015). fine line between pleasure and pain: applying a systems analysis to the Kimberly ultramarathon fire. *6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE. 3*, pp. 1132-1139. Las Vegas, Nevada, USA.: Elsevier Procedia. Retrieved Feb 15th, 2025, from https://www.sciencedirect.com/science/article/pii/S2351978915001900

Herkert J, Borenstein J, Miller K. (2020, July 10th). The Boeing 737 MAX: Lessons for Engineering Ethics. *Science and Engineering Ethics, 26*, 2957–2974. Retrieved August 23rd, 2024, from https://doi.org/10.1007/s11948-020-00252-y

Hitchins, D. (2007). Systems Engineering, A 21st Century Systems Methodology (2007 ed.). West Sussex: John Wiley & Sons Limited. Retrieved April 22nd , 2020, from https://play.google.com/books/reader?id=tdZod1zaIeQC&pg=GBS.PA84

Hudson, P. (2003, Nov 26th). Applying the lessons of high risk industries to health care. BMJ Quality & Safety, 12(suppl1), i7-i12. Retrieved Jan 31st, 2025, from https://doi.org/10.1136/qhc.12.suppl_1.i7

Hutton, R.J.B. & Klein, G. (1999, June). Expert decision making. *Systems Engineering , 2*(1), 32-45. Retrieved Feb 14th, 2025, from https://doi.org/10.1002/(SICI)1520-6858(1999)2:1%3C32::AID-SYS3%3E3.0.CO;2-P

Hutter, B. M., & Jones, C. J. (2006, June). Business risk management practices: The influence of state regulatory agencies and non-state sources. Retrieved Feb 2nd, 2025, from https://www.lse.ac.uk/accounting/assets/CARR/documents/D-P/Disspaper41.pdf

ICAO. (2018). 2.3 Accident causation. In ICAO, the ICAO Safety Management Manual (SMM) (4th ed.). E-BOOK: ICAO. Retrieved September 26th, 2022, from https://www.unitingaviation.com/publications/safetymanagementimplementation/content/#/

ICAO CIRCULAR 247-AW148. (2003, October 6th). *NTSB Docket Item 27 : HUMAN FACTORS DIGEST No. 10 (1993)HUMAN FACTORS. MANAGEMENT AND ORGANIZATION.* Retrieved April 17th, 2023, from NTSB: https://data.ntsb.gov/Docket/?NTSBNumber=CHI01MA006

INCOSE. (2023). International Council on Systems Engineering Handbook, 5th Edition. (D. D. WALDEN, Ed.) Hoboken, NJ 07030, USA: John Wiley & Sons Ltd. Retrieved October 13th, 2023, from https://bit.ly/3IgJeYQ

INCOSE HSI Working Group. (2023). Human Systems Integration , Version 1.2. San Diego,UA: INCOSE. Retrieved December 1st, 2023, from https://portal.incose.org/documents/library

Jack, A Director of Policy, Research and Risk, RSSB. (18 May 2010). *RSSB Response to Appicharla ; Concerns with RSSB Functioning.* RSSB, Policy, Research and Risk. London: RSSB.

Jackson, S. (1997, August). Systems Engineering for Commercial Aircraft. INCOSE International Symposium, 7(1), 36-43. Retrieved Jan 19th, 2025, from https://doi.org/10.1002/j.2334-5837.1997.tb02151.x

Jackson, S. PhD. (2018). Cognitive Bias: A Game Changer for Decision Management?. INSIGHT, 21(4), 41-42. https://doi.org/10.1002/inst.12225

Kahneman.D. (2012). Thinking Fast and Slow. London: Penguin Group.

Kingston, J et al. (2009a, December 20th). The Management Oversight and Risk Tree User Maual and Chart. Retrieved May 7th, 2022, from https://bit.ly/3vTTWzi

Koopman, P., et al. (2019). A safety standard approach for fully autonomous vehicles. Computer Safety, Reliability, and Security: SAFECOMP 2019 Workshops, ASSURE, DECSoS, SASSUR, STRIVE, and WAISE (pp. 326-332)). Turku, Finland: Springer International Publishing. Retrieved February 28th, 2024, from https://bit.ly/3VkcVQb

Langdalen, H., et al. (2020). On the importance of systems thinking when using the ALARP principle for risk management. Reliability Engineering & System Safety, 204, 8. Retrieved April 21st, 2021, from https://bit.ly/3Lvr4oR

Larouzee, J, & Le Coze, J. C. (2020, June). Good and bad reasons: The Swiss cheese model and its critics. Safety Science, 126(June), 11. Retrieved March 22nd, 2023, from https://www.sciencedirect.com/science/article/pii/S0925753520300576

Leggett,L. (2020, Jan 27th). *What went wrong inside Boeing's cockpit?* Retrieved Feb 16th, 2025, from https://www.bbc.co.uk/news/extra/IFtb42kkNv/boeing-two-deadly-crashes#Boeing-e2FBSKSez3

Leveson.N.G. (2011). Engineering a safer world: Systems thinking applied to safety. Retrieved February 23rd, 2024, from https://library.oapen.org/: https://library.oapen.org/handle/20.500.12657/26043

Lopes R.R. (2024, May). *Limitations of Commercial Aviation Safety Assessment Standards Uncovered in the Wake of the Boeing 737 MAX Accidents.* Retrieved Jan 19th, 2025, from https://dspace.mit.edu/handle/1721.1/155396

Macrae, C. (2007, , May). Interrogating the unknown: risk analysis and sensemaking in airline safety oversight. Retrieved Feb 2nd, 2025, from https://www.lse.ac.uk/accounting/assets/CARR/documents/D-P/Disspaper43.pdf

McDermott, T.A., Folds, D.J. and Hallo, L. (2020, July). Addressing Cognitive Bias in Systems Engineering Teams. *INCOSE International Symposium, 30*(1), 257-271. Retrieved May 23rd, 2021, from https://doi.org/10.1002/j.2334-5837.2020.00721.x

Montibeller, G., Von Winterfeldt, D. (2015). Cognitive and Motivational Biases in Decision and Risk Analysis. Risk analysis, 35(7), 1230-1251. Retrieved May 23rd, 2023, from https://onlinelibrary.wiley.com/doi/abs/10.1111/risa.12360

National Transport Safety Committee (KNKT). (2019, October 25th). Aircraft Accident Investigation Report. Retrieved March 8th, 2020, from https://knkt.go.id/: http://knkt.dephub.go.id/knkt/ntsc_aviation/baru/2018%20-%20035%20-%20PK-LQP%20Final%20Report.pdf

Norman, D.A. & Stappers, P.J. (2015). Design X: complex sociotechnical systems. She Ji: The Journal of Design, Economics, and Innovation, 1(2), 83-106. Retrieved Jan 18th, 2025, from https://www.sciencedirect.com/science/article/pii/S240587261530037X?via%3Dihub

NTSB. (2006). Safety Report on the Treatment of Safety-Critical Systems in Transport Airplane, NTSB/SR-06/02. Washington, D.C. 20594: National Transportation Safety Board. Retrieved Novermber 27th, 2024, from https://www.ntsb.gov/safety/safety-studies/Documents/SR0602.pdf

NTSB. (2019a, August 21). SYSTEM SAFETY AND CERTIFICATION SPECIALIST'S REPORT NTSB ID No.: DCA19RA017. WASHINGTON, D.C. 20594: The National Transportation Safety Board. Retrieved August 20th, 2024, from https://downloads.regulations.gov/FAA-2024-0159-0002/attachment_4.pdf

NTSB. (2019b, September 19th). 2019 Safety Recommendation Report 19_ 01 : Assumptions Used in the Safety Assessment Process and the Effects of Multiple Alerts and Indications on Pilot Performance. Retrieved March 8th, 2020, from https://www.ntsb.gov/investigations/AccidentReports/Reports/ASR1901.pdf

NTSB. (2021, March 6th). Investigation of Lion Air Flight 610 and Ethiopian Airlines Flight 302. Retrieved November 9th, 2024, from https://www.ntsb.gov/investigations/Pages/DCA19RA017-DCA19RA101.aspx

NTSB. (2023, January 13th). Response to Final Aircraft Accident Investigation Report Ethiopian Airlines Flight 302 Boeing 737-8 MAX, ET-AVJ. Retrieved January 23rd, 2023, from https://www.ntsb.gov/investigations/Documents/Response%20to%20EAIB%20final%20report.pdf

Oliver, N et al. (2017, Septmber 15th). *The Tragic Crash of Flight AF447 Shows the Unlikely but Catastrophic Consequences of Automation.* Retrieved September 14th, 2021, from Harvard Business Review: https://bit.ly/38qNCpW

Price, H. (2024, May). Big bang, low bar— risk assessment in the public arena. Royal. Society Open Science, 11(5), 1-7. Retrieved Feb 2nd, 2025, from https://royalsocietypublishing.org/doi/pdf/10.1098/rsos.231583

RAIB. (2025, Feb 4th). Re: Non-technical skills in rail accidents: Panacea or pariah? Sixth International Human Factors Rail Conference, 6-9 November 2017, London. Derrby, UK. Retrieved from Yahoo email service.

Rasmussen, J., & Svedung, I. (2000). Proactive Risk Management in a Dynamic Society. Karlstad,Sweden: Swedish Rescue Services Agency. Retrieved February 25th, 2024, from Swedish Rescue Services Agency: https://bit.ly/48wR0bX

Rasmussen,J., Pejtersen, A.M, Goodstein, L.P. (1994). Cognitive Systems Engineering. (A.P.Sage, Ed.) New York: John Wiley and Sons, Inc.

Rasmussen.J. (1997). Risk Management in a Dynamic Society: a modelling problem. Safety Science, 27(2), 183-213. Retrieved 2010, from http://sunnyday.mit.edu/16.863/rasmussen-safetyscience.pdf

Reason, J. (1990a, Apr 12th). The contribution of latent human failures to the breakdown of complex systems. Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 327(1241), 475-484.

Reason,J., et al. (2006, October 30th). Revisting the « Swiss Cheese » Model of Accidents. Retrieved September 23rd, 2011, from https://bit.ly/3FaPU7T

Reason.J. (1990b). Human Error (17th ed.). New York, USA: Cambridge University Press.

Reiman, T. & Rollenhagen, C. (2011, October). Human and organizational biases affecting the management of safety. Reliability Engineering & System Safety, 96(10), 1263-1274. https://www.sciencedirect.com/science/article/abs/pii/S0951832011001086

RSSB. (2020, June). *Development of a new Safety Risk Model.* Retrieved January 21st, 2022, from https://bit.ly/3w3lWAK

Senge. P. (1990b). The fifth discipline: The art and practice of the learning organization. (2006 ed.). New York: Double Day. Retrieved February 15th, 2023, from https://archive.org/details/fifthdisciplineasen00seng/mode/2up

Siemieniuch, C. E., & Sinclair, M. A. (2014). Extending systems ergonomics thinking to accommodate the socio-technical issues of Systems of Systems. Applied ergonomics,, 45(1), 85-98. Retrieved November 5th, 2024, from https://www.sciencedirect.com/science/article/abs/pii/S0003687013000586

Smith, B. E. (2013). NASA/TM—2013–216521 ; Prospective Safety Analysis and the Complex Aviation System. Moffett Field, California: NASA Ames Research Center. Retrieved Jan 17th, 2025, from https://ntrs.nasa.gov/citations/20140010408

Snow,M. Ph. &,Mumaw, Ph.D. (2015). *Preventing Loss of Control in Flight.* Retrieved November 19th, 2024, from https://skybrary.aero/sites/default/files/bookshelf/4361.pdf

Starbuck, W., & Farjoun, M. (Eds.). (2005). Organisation At The Limit: Lessons from Columbian disaster. Oxford, UK: Blackwell Publishing. Retrieved Mau 16th, 2023, from https://www.academia.edu/10718412/Organization_at_the_Limit_Lessons_from_the_Columbia_Disaster

Teal. M Chief Project Engineer, 737 MAX. (2014). New 737 MAX: Improved Fuel Efficiency and Performance. Retrieved November 18th, 2024, from https://skybrary.aero/articles/aero-boeing

The UK HSE. (2001). Reducing Risk Protecting People. Retrieved June 12th, 2021, from https://www.hse.gov.uk/enforce/expert/r2p2.htm

The UK HSE. (2003). Chemical Information Sheet No CHIS7 : Organisational change and major accident hazards. Retrieved December 24th, 2024, from https://www.hse.gov.uk/pubns/chis7.pdf

Tuccio W .A PhD. (2011, Spring). Heuristics to Improve Human Factors Performance in Aviation. Retrieved December 14th, 2021, from The Journal of Aviation/Aerospace Education & Research (JAAER: https://bit.ly/3q6R9ih

Tversky, A.,& Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases- Biases in judgments reveal some heuristics of thinking under uncertainty. Science,New Series, 185(4157), 1124-1131. Retrieved June 11th, 2023, from https://www2.psych.ubc.ca/~schaller/Psyc590Readings/TverskyKahneman1974.pdf

van Kampen, J., et al. (2017, April). Assessing the statistical properties and underlying model structure of fifteen safety constructs. Safety Science, 94, 208-218. Retrieved Jan 31st, 2025, from Safety Science: https://doi.org/10.1016/j.ssci.2016.10.018

Vicente, K. J. (1999). Cognitive work analysis: Toward safe, productive, and healthy computer-based work. CRC Press: CRC Press. Retrieved December 28th, 2024, from https://www.taylorfrancis.com/books/mono/10.1201/b12457/cognitive-work-analysis-kim-vicente

Vicente, K.J. (1998). Human factors and global problems: A systems approach. Systems Engineering, 1(1), 57-69. Retrieved May 16th, 2024, from https://incose.onlinelibrary.wiley.com/doi/10.1002/(SICI)1520-6858(1998)1:1%3C57::AID-SYS6%3E3.0.CO;2-8

Young M.S.;Steel T,RAIB. (2017). Non-technical skills in rail accidents: Panacea or pariah? (pp. 1-10). London: RSSB. Retrieved Feb 15th, 2025, from https://www.rssb.co.uk/spark/sparkitem/pb02612

Zhang, S. et al. (2018). "Analysis on factors of subway incidents for signal system maintenance improving based on a hybrid model.". In A. B. Stein Haugen (Ed.), The 28th annual European Safety and Reliability Conference (ESREL) : Safety and Reliability–Safe Societies in a Changing World, (pp. 17-25). London: CRC Press, Taylor & Francis. Retrieved March 31st, 2023, from https://play.google.com/books/reader?id=vMfADwAAQBAJ&pg=GBS.PA17

Understanding the effect of team familiarity on Shared Spatial Situational Awareness

Vicky Veal & Gulsum Kubra Kaya

Cranfield University, UK

SUMMARY

The study investigates the relationship between team familiarity (both professional and personal) and Shared Spatial Situational Awareness (SSSA) during a flight simulation task. The findings suggest that team familiarity, and specifically personal familiarity, is important for both SSSA accuracy and the percentage of unknown lost SSSA state instances. Interestingly, professional familiarity was not found to be significant.

KEYWORDS

Teams, Situational Awareness, Aviation

Introduction

Situational Awareness (SA) is being aware of your surroundings, interpreting information and anticipating what it means for future states. SA is one of the most critical human factors topics in aviation safety. Indeed, Kharoufah et al. (2018) and Kalagher et al. (2021) analysed numerous aviation accidents and found SA to be the most significant contributory factor. Kalagher et al. (2021) who reviewed 94 general aviation accidents for which situational awareness was cited as being an antecedent, found that the consequence was most often fatal for accidents occurring during the cruise phase in which the crew also had spatial or geographical disorientation. The maintenance of SA is critical for effective decision-making in many domains.

The loss of SA is a problem not exclusive to aerospace, with the loss of SA being identified as a considerable antecedent in many industries' accidents, such as aviation (Kalagher et al., 2021), medical (Al-Moteri, 2022) and maritime (Gommosani et al., 2021).

Much of the accident data attributing a cause of SA tends to relate to teams rather than individuals, such as a flight deck crew, operating theatre, or a control room. A review of the literature however demonstrates a vastly reduced quantity of research relating to team or shared SA, which may therefore fail to acknowledge that cockpits and control rooms are made up of more than one individual; it fails to consider SA within the team (Graafland et al., 2014). To this end, this research investigates the relationship between team familiarity and Shared Situational Awareness (SSA). To evaluate and measure this, the research specifically focuses on the spatial element of SSA, herein identified as Shared Spatial Situational Awareness (SSSA).

Situational Awareness

There are many definitions and theories of SA, including that by Smith and Hancock (1995), but perhaps the most prevalent definition is that by Endsley (1995), which defines SA as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future". Endsley's model suggests three hierarchical stages: level 1 SA - perception of the elements in the environment, level 2 SA – comprehension of the current situation, and level 3 SA – projection of future status.

There are many factors which influence SA according to Endsley (1999), including processing limitations such as attention and working memory limitations, as well as coping mechanisms such as mental models and automaticity. Mental models and schema for example are used as coping mechanisms to allow pilots to speed up decision-making in the higher levels of SA.

Endsley (1995) further suggests that there can be sub-categories of SA, such as spatial or geographical SA, which includes requirements across each of the three levels of SA of her model. Her paper goes on to state that SA has often been observed to be highly spatial, and that this spatial element is valuable in establishing the importance of elements of the environment, and therefore which elements to prioritise at a given time. For this reason, this paper focuses on spatial/geographical SA, referred to as 'spatial' from this point forward.

Shared Situational Awareness

Endsley (1995) defines Team SA as "the degree to which every team member possesses the SA required for his or her responsibilities". She and Li (2017) conducted a literature review which included Ensley's work and summarised the definition of team SA as "a multidimensional construct" which "contains the SA related to the individual's own roles, the SA of other team members, and the SA of the overall team." It is important to understand that these definitions support the notion that there are elements of a person's SA within a team that are shared and elements which are not. It is only necessary for a subset of information to be shared across a team (Endsley, 1995). Shared Situational Awareness (SSA) is therefore a sub-component of team SA and is defined by Endsley and Jones (1997) as "the degree to which team members possess the same SA on shared SA requirements." Endsley (1995) suggests it is the overlapped elements (shared SA), which require a large proportion of the team's coordination, which could be done via many means including verbal communication or duplicated display information.

Endsley (1999) suggests to develop team SA, much of the team's coordination activities are based on information-sharing. This will include level 1 SA information such as data from visual displays, as well as information pertaining to levels 2 and 3 so that the team can benefit from the varying experiences of the individual team members. It is less likely that the higher levels of SA will be achieved via displays, making team SA more important to achieve levels 2 and 3. This may for example be done via communication and a Shared Mental Model (SMM) between team members.

The Role of Shared Mental Models and Communication

Shared Mental Models and communication elevate themselves as being key themes within the literature for the achievement of Levels 2 and 3 SA, with Endsley and Jones (2004) stating that higher team performance was found in teams that had SMMs in comparison to those that did not. Indeed, in an analysis of accident data, Jones and Endsley (1996) found 13.8% of the accidents reviewed to be attributable in some way to mental models.

Team members' mental models need to be the same to come to the same conclusion and therefore share the same SSA (Endsley & Jones, 1997). This is supported by Cannon-Bowers et al., 1993, as cited in Endsley, 1999, suggesting that the sharing of SA between the team members is enriched by SMMs "which provide a common frame of reference for crew member actions and allow team members to predict each other's behaviours". Stout et al. (2017) further suggest that SMMs are "a prerequisite for achieving team situational awareness" and are critical for performance.

Endsley (1999) suggests that by having similar mental models, the team are better able to communicate with each other as they can predict each other's actions, therefore enhancing SA.

Orasanu and Salas (1993, as cited in Endsley & Jones, 2004), state that one of the key advantages of SMMs is that "they can provide important interaction patterns and standardized speech and communications allowing crew members to interact in predictable ways".

However, Endsley and Jones (1997) suggest that higher levels can be achieved by the team if they possess the same mental models without the need for extra communication. This is supported by Mosier and Chidester (1991, as cited in Endsley, 1999) who found there to be less communication between better-performing teams.

Shared Situational Awareness States

Endsley and Jones (2004) define team SSA states into three categories; both correct, both incorrect, and different. If both team members have the same understanding and are correct, this is the most preferred state. Alternatively, they may have differing views with either only one of the team being correct, or them both being incorrect (in different ways or both having the same incorrect understanding).

Losing SA however, does not necessarily lead to poor performance and therefore risk of danger. In a tactical fighter mission study, Endsley (1990) found that if they realised they had lost SA it was not necessarily linked to poor performance when they could modify their behaviour. If the crew do not know they have lost SA, i.e. neither pilot has an understanding that SA has been lost, it carries the greatest safety risk (Jentsch et al., 1997, as cited in Endsley & Jones, 1997). To this end, the method used when understanding unknown lost SSSA will focus on occurrences where both participants are unaware rather than only one.

Shared Spatial Situational Awareness

There is little research available specifically looking at the spatial element of SSA. One such study by Prebot et al. (2019), investigated the assessment of SSA within complex decision-making via the assessment of the spatial element of SSA. The experiment provided verification of the method used to test SSSA, stating that "using distances (spatiality) as an evaluation metric of Shared SA, allowed us to quantify more precisely the situation awareness, giving levels to in the end qualify teams performance." For this reason, a similar method will be used in our research.

Familiarity

Familiarity has been defined as a multidimensional construct consisting of both professional and personal familiarity by Hanft (2002, as cited in Maynard et al., 2018). There have been many studies conducted that show team familiarity to have a positive outcome on performance, despite Barker et al. (1996) claiming no difference. A study by Kurmann et al. (2014) compared surgical outcomes of patients who had been operated on by newly formed teams, against patient outcomes when the teams were in the last month of a six-month partnership. They found improved team performance, including higher concentration scores, and reduced morbidity in the teams that had been partnered for longer, suggesting team familiarity to be beneficial. Muskat et al. (2022), who conducted a systematic literature review on team familiarity across many domains, supported this, finding that team familiarity resulted in positive performance, higher team cognition, higher efficiency, increased output quality, increased team productivity, error reduction and higher safety.

The National Transportation Safety Board (NTSB) (1994) conducted a review of major accidents and incidents of U.S. air carriers from 1978-1990 that involved flight crew and found 73% of the incidents occurred on the first day of a pairing, with 44% being on the crew's first flight. Similarly, a review of U.S. air carrier accident data between 1991 and 2010 by Boss et al. (2013) looked at the accident rates of Captains and First officers with team familiarity, including measuring this via whether they were on their first day of the pairing. They found a "negative effect on accident rates" with higher accident rates amongst those with lower crew familiarity. This is supported by Endsley and Jones (2004) who suggest that teams that have not previously worked together might be at a disadvantage, as those that have, are able to understand each other, their roles etc, which in turn helps promote effective communication and anticipate actions. This is also supported by Orasanu and Salas (1993, as cited in Endsley, 1997) who suggest that the familiarity between the crew and shared experiences allow the development of the SMMs. This in turn is important in allowing the crew to reach levels 2 and 3 SA (Endsley, 1995), and therefore allowing good team performance.

The Measurement of Team Familiarity

The systematic review by Muskat et al. (2022) suggests that there are several antecedents to team familiarity, with the key antecedent being cited as time. This includes the amount of time working together, the number of shared experiences, and the amount of friendship. They do not however state the amount of time necessary to be categorised as a familiar team, or the types of activities, depth etc of the experiences to achieve this.

Maynard et al. (2018) measured personal and professional familiarity using a Likert scale questionnaire based on the work by Hanft (2002, as cited in Maynard et al., 2018). They found that professional familiarity was significantly positively related to their task of information elaboration but that personal familiarity was not. In turn, information elaboration was significantly related to team effectiveness.

Professional and Personal Team Familiarity

There is little research available with respect to how professional and personal team familiarity might affect SA; however, theoretical underpinnings suggest a difference.

When considering professional familiarity, SMMs and Endsley's SA theory (1995) would suggest that shared work experiences allow the building of professional information, such as an understanding of the knowledge bases of teammates and the expertise that can be relied upon. This would allow for more effective communications and therefore anticipation of each other's actions (SMM). This in turn allows higher levels of SA to be achieved within the team (Endsley & Jones, 2004). This is supported by Roberston and Endsley (1995, as cited in Endsley, 1999) who suggest that Crew Resource Management (CRM) training can impact the development of SMMs and individual SA.

There is however far less written about how personal familiarity might lead to better performance. One explanation however is offered by Maynard et al. (2018), who suggest that increased familiarity means that members "are more likely to suggest and use the ideas of others based on their levels of interpersonal comfort, which reduces fear of ridicule or exclusion". Communication and information elaboration are social phenomena requiring the free exchange of information, integration, and decision-making. Edmondson (1999) suggests that this is more likely if there is psychological safety and comfort with their teammates. As more effective communication is suggested to be linked to better SMMs, it could be suggested that personal familiarity would result in better SSSA. This is supported by Jehn and Shah (1997), who found significantly better performance from friendship groups compared to acquaintance groups regarding decision-making due to greater commitment and cooperation.

Method

This study investigates the relationship between team familiarity (professional and personal) and SSSA. It was hypothesised that there would be a difference in the team's SSSA accuracy based on team familiarity, as well as a difference in the 'unknown lost SSSA state' shown, where the teams incorrectly think that they are marking the same location as their teammate. This study conducted an experiment with 74 participants, and the study design includes pre-experiment, experiment and post-experiment phases.

In the pre-experiment phase, participants were voluntarily assigned to either the navigator or pilot role at the start of each session after collecting informed consent. Then participants' subjective evaluation of team familiarity, including personal and professional elements, was recorded via a seven-point Likert scale (based on Hanft, 2002, and Maynard et al., 2018). The Qualtrics questionnaire also captured demographic and background information. The experiment was then explained and instruments (for example, a map and recording sheet, as well as the target for the navigator) were introduced to the participants. The pilot then had a practice flight for three minutes, whilst the navigator was invited to watch for the practice only to gain an understanding of the visuals their teammate would be experiencing.

In the experiment phase, the pilot performed an online flight simulation game based on Bing maps (provided by geo-fs.com (Geo-FS, 2023) in HD) using a Thrustmaster USB joystick set with a 0.3 sensitivity, whilst the navigator verbally guided the pilot to a target known only to themselves (Hayward executive airport). They were allowed to verbally communicate freely to ascertain location, but the navigator was instructed not to directly give the target location (for example, 'top right of the map'). The participants had no previous knowledge of where they would be flying (San Francisco) and were separated by a barrier so that no line of sight was possible. Each participant was given an A3 piece of paper with a map of the San Francisco area printed on it, taken from Bing maps. All participants started at the same point and were already in the air. The starting location and direction were indicated on the map, with due North at the top of the paper as convention would dictate. Participants could choose any route. No coordinates, grid or scale indications were given.

Every 45 seconds, the simulation would be paused, and participants were asked to mark down where they thought the aircraft was located on their paper map. This is referred to as the Position Evaluation Point herein (PEP). They were also asked whether they thought they had marked the same approximate location as their teammate. The participants were asked not to communicate during this time, and the actual location was recorded. This was repeated until the aircraft landed or until they had been flying for 20 minutes, at which point the simulation was ended.

During the experiment phase, SSSA accuracy was recorded by measuring the distance between the PEP and the actual aircraft location for each participant (mm). The actual location was ascertained from a screen recording of the flight. In addition, team awareness of SSSA state was measured when asked if they had marked the same approximate location as their teammate (yes/no). The percentage of instances where both participants responded 'yes' but were more than 10mm apart determined the unknown lost SSSA score.

In the post-experiment phase, participants were debriefed and thanked for their participation.

Results

74 participants (52 Male, 20 Female and two declined to answer) ranging from 18 to 69 years old (M= 33.08) were recruited on an availability basis at their workplace to complete this study. Among the 37 tested teams, 10 did not reach the target. Six teams landed at an incorrect location and four failed to land within 20 minutes. Of the successful teams, three sets had incomplete data, and so the data relating to these were removed, along with those that did not reach the target.

The findings suggest that overall team familiarity ($p = -.478$, $p < 0.05$) and especially personal familiarity ($r = -.499$, $p < 0.05$) are important for SSSA accuracy. It was also found that overall team familiarity ($p = -.424$, $p < 0.05$) and especially personal familiarity ($r = -.473$, $p < 0.05$) are important for the percentage of unknown lost SSSA state instances. This suggests that higher overall team familiarity and personal familiarity scores are related to lower team accuracy scores, i.e., more accurate results within the team and a lower number of instances of unknown lost SSSA. Interestingly, professional familiarity was not significantly correlated.

Discussion

The relationship between team familiarity (professional and personal) and SSSA has been explored using a correlational study design in which teams directed each other to a target in an online flight simulation game to understand their SSSA accuracy and SSSA state.

The results of this study confirmed prior theory and research indicating there to be a positive relationship between team familiarity and performance. Whilst there was little research available to consider the effect of SSSA specifically, it could be hypothesised that there would be an effect on SSSA accuracy, as well as instances of unknown lost SSSA states, as a result of overall team familiarity, professional familiarity, and personal familiarity based on Endsley's work of SA and team SA/shared SA (1995). The literature suggests that team familiarity was found to be positively correlated with shared experiences, with an emphasis on work-related experiences, which allow team members to build an accurate SMM and therefore communicate more effectively (Endsley, 1999). This in turn allows them to act in more predictable ways (Orasanu & Salas, 1993, as cited in Endsley & Jones, 2004). In addition, the literature suggests that personal familiarity is positively correlated as communication is enhanced by friendship and personal familiarity due in part to the more likely free exchange of information (Edmondson, 1999). One participant stated, "Because we knew each other, I could tell when she wasn't sure or if something was going wrong because of her tone of voice". This would support shared experiences being important for effective communication within the team, as suggested to be important for the development of SMMs and therefore reaching levels 2 and 3 SA by Endsley (1999).

There was however a lack of support with regard to the relationship specifically between professional team familiarity and SSSA with respect to both accuracy and unknown lost SSSA states. This was in contrast, however, to the work by Maynard et al. (2018) who found that professional rather than personal familiarity led to better information elaboration. One possible explanation for this is related to the fact that participants weren't familiar with this task type. Endsley (1999) states that novice pilots have limitations due to working memory limitations. The cockpit is a complex environment that can exceed attention and working memory capacities, suggesting that SA can be lost if the information that needs to be perceived isn't attended to whilst attending to other sources. Whilst participants were given practice time in the pre-experiment phase, they were still novices with regard to the task type. The practice session would not have been enough to reduce the demands on their attention significantly. This would support the need for further assessment with a representative flight deck crew and a representative task, which was not possible here due to the limitations with regard to resources on a student project. However, this study provides valuable insight into the relationship as it is one of the first studies to consider the relationship between SSSA and team familiarity, but it is acknowledged that the findings should be validated by flight crew and representative flight tasks.

Key Takeaways/Applications

There are numerous practical implications that can be proposed due to this research. As the results suggest personal familiarity is beneficial when considering SSSA, managers may want to consider adding personal familiarity as an element of their decision-making criteria when deciding on crew rosters. Whilst there will be many other factors to be considered with varying levels of importance and risk, such as working time regulations, if there is existent personal familiarity between potential crews, it may be of benefit to roster them together as this could provide less risk relating to SSSA as a result.

If it is not possible to take advantage of existent personal familiarity when rostering, it may be beneficial to take steps to enhance the organisational culture to encourage personal familiarity within their staffing. This may for example be via team-building days that do not solely focus on work-related tasks or encouraging social interactions such as sports clubs and team lunches. Indeed, Maynard and Gilson (2021) suggested several methods to improve personal familiarity, including arriving at meetings early to initiate organic personal conversations. This could easily be adopted by flight deck crew arriving slightly early rather than getting straight down to business.

Additionally, although not as much weight should be given compared to personal familiarity, it would be worth using existing professional familiarity as a selection criterion when considering rosters. This is because team familiarity, which included both personal and professional elements, was found to be significant. Again, if not already existing, techniques could be employed to enhance it.

It is recommended that awareness be raised on the benefits of team familiarity, especially personal familiarity, for any high-risk industry in which SSSA is critical. The implementation of specific training to counter this may be of use.

Limitations

Based on the literature review conducted, this is one of the first studies to consider the relationship between SSSA and team familiarity, even without further consideration for the sub-type of familiarity. Whilst the results suggest the saliency of personal familiarity, future research is needed to address some research limitations.

This research was conducted in a laboratory setting with a sample of volunteer participants who were not pilots. Future research should aim to validate findings using pilots in a representative environment, which was impossible due to limited access to resources on a student project. The correlational design of this research has allowed the identification of a potential relationship between team familiarity and SSSA. However, an experimental design should be adopted for follow-up investigations to establish causation and build upon this evidence. Anecdotally, this difference in crew team familiarity already naturally exists within the differences between commercial and military operations. Military crews are more likely to be flying with familiar teammates as there are generally a smaller number of pilots trained to fly each aircraft type hence higher levels of professional familiarity, but they are also more likely to know each other personally due to living on military bases. Accurate SSSA as well as accurately perceived SSSA is extremely important when flying to avoid the likes of Controlled Flight Into Terrain and Mid-Air Collisions, but it is even more important for military aircraft in potential combat situations due to increased dangers. It is unfortunate that SSSA cannot be isolated and naturally measured, for example via accident data analysis, as a result of the existing differences. Due to many interacting variables, such as the level of training received and the nature of the tasks being undertaken, SSSA accident data relating to team familiarity would be difficult to decipher as being purely a result of team familiarity. To that end, experimental data should be gathered to explore the correlation further and validate findings.

Future work could be conducted to understand whether there was anything significant about the data relating to teams that did not successfully complete the task. It may be that SSSA contributed to these teams being unsuccessful in completing the task, suggesting that any team familiarity effect would be even more informative.

Additionally, practical implications such as those made above, should not be considered in isolation. Whilst team familiarity has been shown to positively affect SSSA within this research, it does not consider any other variables that could potentially have a negative effect. If a third variable was negatively affected by team familiarity, a safety management systems approach might be conducive to understanding the risk associated with any proposed approach to ensure the safest route is taken for operations. Future research might consider further understanding performance factors, including workload, in terms of personal and professional familiarity to aid in quantitatively assessing appropriate pairings for rosters.

A final limitation of this research is that it only considered a two-person cockpit team. In reality, flying aircraft involves many people in different roles as acknowledged by ICAO who identify Air Traffic Controllers, cabin crew and aircraft dispatchers as "essential participants in an effective CRM process" (ICAO, n.d.). To this end, future research should consider whether there is also an impact on SSSA due to team familiarities between these extended teams, such as between the flight deck crew and ATCs, which would allow the implementation of practical recommendations to be benefited from more widely. Certainly, Maynard et al. (2018) and Maynard and Gilson (2021) found that information elaboration is related to professional familiarity more saliently than personal in virtual teams, with virtuality moderating the relationship. The relationship between the flight deck and ATC may be considered 'virtual' and therefore potentially result in a different impact on SSSA compared to a face-to-face relationship.

Conclusion

Whilst much of the literature suggests that shared work and task-related experiences are important to build role understanding and a knowledge base of teammates (professional familiarity) in order to gain a better mental model and therefore better SA, it seems that this research found the personal element of shared experiences to be more pertinent to the development of SSSA. This element is seldom discussed in the literature to date. It is hoped that this research can bridge the gap in the literature relating to personal, professional, and overall team familiarity, as well as SSSA.

References

Al-Moteri, M., Alfuraydi, A. A., Alsawat, A. Z., Almulhis, R. S., Alnadwi, B. S., Youssef, H. A., & Althobiti, E. S. (2022). Shared situational awareness within the Hospital Emergency Context: A scoping review. *Healthcare*, *10*(8), 1542. https://doi.org/10.3390/healthcare10081542

Barker, J. M., Clothier, C. C., Woody, J. R., McKinney, E. H., & Brown, J. L. (1996) Crew resource management: a simulator study comparing fixed versus formed aircrews. *Aviation, Space, Environmental Medicine, 67*(1), 3-7.

Boss, K. K., Depperschmidt, C. L., Mwavita, M., & Bliss, T. J. (2013). Characteristics of pilots involved in U.S. air carrier accidents between 1991 and 2010. *Collegiate Aviation Review International*, *32*(1), 14–38. https://doi.org/10.22488/okstate.18.100510

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*(2), 350–383. https://doi.org/10.2307/2666999

Endsley, M. R. (1990). Predictive Utility of an Objective Measure of Situation Awareness. Proceedings of the Human Factors Society Annual Meeting, 34(1), 41-45. https://doi.org/10.1177/154193129003400110

Endsley, M. R. (1995). Toward a theory of situation awareness in Dynamic Systems. Human Factors: The Journal of the Human Factors and Ergonomics Society, 37(1), 32–64. https://doi.org/10.1518/001872095779049543

Endsley, M. R. (1999). Situation awareness in aviation systems. In D. J. Garland, J. A. Wise, & V. D. Hopkin (Eds.), Handbook of aviation human factors (pp. 257–276). Lawrence Erlbaum Associates Publishers.

Endsley, M. R., & Jones, W. M. (1997). Situation awareness, information dominance, and information warfare. Wright-Patterson AFB, OH: United States Air Force Armstrong Laboratory.

Endsley, M. R., & Jones, D. G. (2004). Designing for situation awareness: An approach to user-centered design (2nd ed.). CRC Press.
Geo-FS. (2023). *The accessible flight simulator*. GeoFS - Free Online Flight Simulator. https://www.geo-fs.com/ 35
Gommosani, M., Turan, O., & Kurt, R. (2021, June). Analysis of maritime accidents due to poor situational awareness. In *1st International Conference on the Stability and Safety of Ships and Ocean Vehicles*.
Graafland, M., Schraagen, J. M., Boermeester, M. A., Bemelman, W. A., & Schijven, M. P. (2014). Training situational awareness to reduce surgical errors in the operating room. *British Journal of Surgery*, *102*(1), 16–23. https://doi.org/10.1002/bjs.9643
Hanft, T. R. (2002). Familiarity in organizations. [PhD thesis, The University of Texas at Dallas]. https://www.proquest.com/openview/39349ec19fd29caf427fbdb5b228f603/1?pq-origsite=gscholar&cbl=18750&diss=y
ICAO. (n.d.). *Crew Resource Management Training Programme*. National Civil Aviation Administration. https://www.icao.int/APAC/RASG/SafetyTools/04%20Advisory%20Circular%20%E2%80%94%20Crew%20Resource%20Management%20Training%20Programme%20CRM.pdf
Jehn, K. A., & Shah, P. P. (1997). Interpersonal relationships and task performance: An examination of mediation processes in friendship and acquaintance groups. *Journal of Personality and Social Psychology*, *72*(4), 775–790. https://doi.org/10.1037/0022-3514.72.4.775
Jones, D. G., & Endsley, M. R. (1996). Sources of situation awareness errors in aviation. *Aviation, space, and environmental medicine*, *67*(6), 507-512.
Kalagher, H., de Voogt, A., & Boulter, C. (2021). Situational Awareness and general aviation accidents. *Aviation Psychology and Applied Human Factors*, *11*(2), 112–117. https://doi.org/10.1027/2192-0923/a000207
Kharoufah, H., Murray, J., Baxter, G., & Wild, G. (2018). A review of Human Factors Causations in commercial air transport accidents and incidents: From to 2000–2016. *Progress in Aerospace Sciences*, *99*, 1–13. https://doi.org/10.1016/j.paerosci.2018.03.002 36
Kurmann, A., Keller, S., Tschan-Semmer, F., Seelandt, J., Semmer, N. K., Candinas, D., & Beldi, G. (2014). Impact of team familiarity in the operating room on surgical complications. *World Journal of Surgery*, *38*(12), 3047–3052. https://doi.org/10.1007/s00268-014-2680-2
Maynard, M. T., & Gilson, L. L. (2021). Getting to know you: The importance of familiarity in virtual teams. *Organizational Dynamics*, *50*(1), 100844. https://doi.org/10.1016/j.orgdyn.2021.100844
Maynard, M. T., Mathieu, J. E., Gilson, L. L., R. Sanchez, D., & Dean, M. D. (2018). Do I really know you and does it matter? unpacking the relationship between familiarity and information elaboration in global virtual teams. *Group & Organization Management*, *44*(1), 3–37. https://doi.org/10.1177/1059601118785842
Muskat, B., Anand, A., Contessotto, C., Tan, A. H., & Park, G. (2022). Team familiarity—boon for routines, Bane for Innovation? A review and future research agenda. *Human Resource Management Review*, *32*(4), 100892. https://doi.org/10.1016/j.hrmr.2021.100892
National Transportation Safety Board. (1994, January). *A review of Flightcrew-involved major accidents of U.S. air carriers 1978 through 1990*. https://www.ntsb.gov/safety/safety-studies/Documents/SS9401.pdf
Prebot, B., Salotti, J.-M., Vennin, C., & Claverie, B. (2019). Shared spatial situation awareness as a team performance indicator in Collaborative Spatial Orientation Task. *Advances in Human Error, Reliability, Resilience, and Performance*, 106–115. https://doi.org/10.1007/978-3-030-20037-4_10

She, M., & Li, Z. (2017). Team situation awareness: A review of definitions and conceptual models. *Engineering Psychology and Cognitive Ergonomics: Performance, Emotion and Situation Awareness*, 406–415. https://doi.org/10.1007/978-3-319-58472-0_31
Smith, K., & Hancock, P. A. (1995). Situation awareness is adaptive, externally directed consciousness. Human factors, 37(1), 137-148.
Stout, R. J., Cannon-Bowers, J. A., & Salas, E. (2017). The role of shared mental models in developing team situational awareness: Implications for training. *Situational Awareness*, 287–318. https://doi.org/10.4324/9781315087924-18 l

Workload and Perceived Usefulness when an Electronic Checklist with Sound is Used for Aeroplane Landing

Florin Dumitrascu

Cranfield University, UK

SUMMARY

This study explored whether there were differences in workload and perceived usefulness between people who used an electronic checklist with sound for aeroplane landing and those who used a paper checklist. Two groups of University students were assigned to one of two conditions (A. paper checklist and B. electronic checklist with sound). Each group read their checklist (paper – group A, electronic checklist – group B) and selected cockpit areas on a screen to indicate completion of landing sub-tasks on an A320 aircraft. Workload and perceived usefulness were assessed subjectively. Those who used the electronic checklist stated lower levels of workload and found their checklist more useful than those who used the paper checklist. These findings suggest that electronic checklists with sound are a promising alternative to paper checklists used for landing, but further research is needed to fully understand their benefits for pilots.

KEYWORDS

Electronic Checklist, Sound, Paper checklist, Workload, Performance, Blue Colour, Cue

Introduction

Paper checklists have been in use for many years in different settings. They are used to guide, provide information and overcome the limitations of short-term and long-term memory. One of the tasks that paper checklists have traditionally been used for by pilots is to help them with landing, with one of the pilots reading the items on the checklist aloud and the other completing the required tasks. This may change in the future with the 'single-pilot' concept, whereby there will only be one pilot in the cockpit (Liu et al., (2016). This could have a detrimental impact on workload during landing because all the information would need to be processed and tasks completed by one person instead of two.

A shift towards integrating electronic checklists in the cockpit has recently been observed on some types of aircrafts (e.g. the Boeing 777 and the Airbus A220). Pilots can use those checklists to complete the landing tasks in a way similar to the paper checklists. Using electronic checklists for landing is a promising method because, in contrast to paper checklists, they can provide feedback to pilots about the read and completed actions. This can reduce the likelihood of omitting procedures, thus help pilots to avoid mental overload during the already mentally demanding stage of landing, and improve their user experience (Boorman, 2001a/b, as cited in Hales & Pronovost, 2006). This study explored if workload in single-pilot operations can be lower and perceived usefulness higher when an electronic checklist with sound is used to complete landing tasks than a paper one.

Method

A between-subject study design was used with 30 Cranfield University students being assigned to one of two conditions (paper – group A, electronic checklist – group B). For the purposes of the study, a PowerPoint presentation was developed to simulate the parts of a A320 cockpit that pilots would be expected to interact with to land the aeroplane. Both groups were requested to read each of the items on their landing checklist and then use the mouse to click on the correct part of the screen to simulate completing the landing sub-tasks. Before using the mouse to click on the laptop screen, Group B also had to select each of the items on their checklist. When an item was correctly selected on the electronic checklist, a sound was heard to reinforce the participants to use the mouse to make a selection on the laptop and the last item that was selected on the checklist was highlighted in blue to reduce search time. The NASA-TLX Scale (workload) and an 8-point Likert item and an open-type question (perceived usefulness) were used.

Results

A series of Mann-Whitney U tests were conducted to compare the NASA-TLX ratings on each of the six subscales between the two groups. The analysis showed that the group who used the electronic checklist with sound reported significantly lower levels of mental demand ($p=.02$), frustration ($p=.02$), and effort ($p=.04$) and better performance ($p=.02$) than the group that used the paper checklist. There were no statistically significant between-group differences in the reported physical ($p< .299$) and temporal demand ($p< .281$).

An independent sample t-test analysis was performed to determine whether there was any difference between the groups in how useful they found their checklists. The mean reported Usefulness score was significantly higher in the group that used the electronic checklist, ($t(29)=-3.163$, $p< .004$, $d=-1.137$). Subjective feedback revealed that the electronic checklist helped the participants track their progress with tasks and provided them with real-time feedback. In contrast, according to the participants, there is a risk that landing tasks/checks are omitted when using a paper checklist.

Discussion

The findings of this research are consistent with past research, arguing that human-centred design of augmented visualisation and auditory aids human-machine interaction performance (Li et al., 2020). This is arguably because the sound should compensate for the attention required for each individual task to be completed. Therefore, reducing the time glancing to the checklist items. This is mainly because attention is a stimuli-driven and is goal-directed that works simultaneously to complete the task (Eysenck & Keane, 2015; Baker et al., 2004).

Several confounding variables could have affected the results of the experiment, this includes usage of naïve participants, that were not pilots; low fidelity replication of A320 interface and checklist, which was designed in a PowerPoint MS; and did not employ eye tracking equipment to assess the level of attention switching between the two groups. Therefore, future research with more complex control might be more suitable to create a realistic interaction and use of eye tracking equipment to measure time required to glaze at each item might reveal whether sound brings any benefits.

Conclusion

In conclusion, this thesis has endeavoured to provide some light on whether there were differences in workload and perceived usefulness between paper checklist and electronic checklist with sound . Electronic checklist with sound has proven to show low levels of workload and it was found more useful than the paper checklist.

References

Baker, K., Esgate, A., Groome, D., Heathcote, D., Kemp, R., Maguire, M., & Reed, C. (2004). *An introduction to applied cognitive psychology*. Psychology Press

Boorman, D. (2001a). Safety benefits of electronic checklists – An analysis of commercial transport accidents. Proceedings of the 11th International Symposium on Aviation Psychology, 1-6; Columbus, OH: The Ohio State University.http://www.flighttestsafety.org/images/Boorman_OSU2001_Paper.pdf.

Boorman, D. (2001b). Today's electronic checklists reduce likelihood of crew errors and help prevent mishaps. ICAO Journal, 56, 17-20, 36. Montreal, Canada: ICAO.

Eysenck, M. W., & Keane, M. T. (2015). *Cognitive psychology: A student's handbook*. Psychology press. https://doi.org/10.4324/9781315778006

Li, W. C., Bord, T., Zhang, J., Braithwaite, G., & Lone, M. (2020). Evaluating system usability of augmented reality in flight operations. *Contemporary ergonomics and human factors.*

Liu, J., Gardi, A., Ramasamy, S., Lim, Y., & Sabatini, R. (2016). Cognitive pilot-aircraft interface for single-pilot operations. *Knowledge-based systems*, *112*, 37-53. https://doi.org/10.1016/j.knosys.2016.08.031

Adaptive safety on the construction frontline

Clinton Horn, Patrick Waterson & Gyuchan Thomas Jun

Human Factors and Complex Systems Group, Loughborough University

SUMMARY

Frontline construction labourers make performance adaptations from the safety rules and the prescribed safe work method statements (SWMS) by adopting their preferred way of working developed from their previous learned experiences. Performance adaptations are motivated by the inherent production pressures within the construction industry driving a need for efficiency, and a constantly changing working environment. There is a need to rethink how safety and safety performance outcomes are constructed on the frontline of construction.

KEYWORDS

Adaptive safety, Construction, Job and work design

Introduction

Traditional safety thinking would suggest that normative safety artefacts such as safety rules, work instructions, procedures and risk assessments are safe in and of themselves, and that adverse events are simply as a result of workers deviating from them (Woods and Hollnagel, 2006). Contrastingly, the adaptive safety literature argues that such artefacts, no matter how well intended, will inevitably be under-specified as “there will always be contextualised couplings that are unknown” (Hollnagel, 2018, p.10). In such situations, workers will draw on their learned capacities to adapt to such intractableness (Reiman et al., 2015), with safety performance outcomes, whether adverse or otherwise, being an emergent property of this construct.

However, such assertions are in need of ‘reality-based’ empirical investigations (Rae et al., 2020) that are “grounded in operational reality” (McDonald, 2006, p. 180) necessary to evidence and understand the underlying sociotechnical constructs and their manifestation. This study contributes to this evidence gap by empirically investigating frontline construction labourer attitudes towards those normative safety artefacts and how they interact with a range of sociotechnical constructs to influence why and how construction labourers adapt their performance if not work-to-rule.

Additionally, production pressure is a ubiquitous characteristic and constraint within the construction industry (e.g. Haslam et al., 2005; Han et al., 2014). However, how this upstream-derived constraint manifests further downstream is less well understood. This study also contributes to the literature in this regard.

Methodology

This was a qualitative study using 1:1 semi-structured interviews involving 20 frontline construction labourer participants working within the ground engineering sector of the UK construction industry. The interview transcripts were thematically coded using Nvivo14 and analysed using a reflexive thematic analysis (RTA) methodology (Braun and Clarke, 2020).

Findings

Participants acknowledged that the safety rules are intended to *"keep everyone safe"*. Although, the extent of compliance varies. For example, *"they should be followed 100%, but it is not always possible to follow them"*.

The findings revealed that there are three primary drivers motivating labourer performance adaptations, 1) the inherent production pressures within the construction industry. For example, *"do anything you have to do to get the job done"*, 2) a constantly changing working environment. For example, *"we should have used a forklift to move all these steel plates, but we [physically] couldn't get the forklift past the digger because [the gap] was very narrow",* and, 3), the social dynamics shaping the relationship between labourers and their supervisors which was found to have a significant mediating effect on labourers either working-to-rule or deviating from the safety rules by adapting their performance. For example, "*whenever you hear about [labourers] bending the rules in the method statement or not doing something that it [prescribes] then, yes, that is what it probably comes down to"*.

The findings suggest that how labourers adapt their performance is informed largely by their previous learned experience, resulting in labourers having developed a preferred way of performing a task. If the prescribed methodology mandated within the normative safety artefacts closely aligns with their preferred way of working, then the likelihood for labourer adaptions is significantly reduced. However, in protecting their preferred ways of working, that is also developed to ensure efficiency in performance, the more experienced amongst the labour crews will often take it upon themselves to 'take under their wing' the less experienced and pass on their experience and knowledge of the task. Such informal strategies also develop trust (in performance capabilities) amongst labour crews which was found to be important amongst the participants.

Linked to the extent of labourers' previous learned experience informing how they adapt, or if they instead choose to work-to-rule, the study found that a labourer's attitude to risk, in the context of rule-breaking, was closely aligned with the extent to which they were comfortable deviating from the safety rules and adapting their performance, including whether the adaptation would enhance or compromise safety. Performance adaptations that risk compromising safety was found to be a 'red line' for all participants.

Discussion and key takeaways

This study has revealed that construction labourers have developed a preferred way of working that is inextricably linked with a desire to perform tasks efficiently whilst also maintaining safety. This preferred way of working is constantly evolving as the workers' learned experience of the task evolves, meaning that they develop more and more reference points to draw upon when performing the task. It also highlights how safety can potentially be compromised amongst those labourers whose previous learned experience is immature and weak relative to other more experienced labourers who have more mature reference points to draw upon when adapting their performance and, thereby, increasing the likelihood for "positive schema transfers" (Tversky and Kahneman, 1974, p.1128).

The effect being that, with the construction project environment being a temporary and transient reality (Davies, 2017), these varying levels of learned experience transiting the construction frontline will either contribute to organisational resilience, including safety, or compromise it. As a result, safety performance outcomes, whether adverse or otherwise, emerges "through the co-evolution of sociotechnical structure and [worker] agency" (Furniss et al., 2019, p.687).

With construction contractor organisations largely blind to the manifestation of these sociotechnical constructs, there is an opportunity for the findings of this study, and the currently under-way follow-on studies, to develop operationally practical methods and tools specific to the construction industry that organisations can use to both surface and leverage these constructs. In an industry where there are consistently more work-related fatalities than any other industry (HSE, 2024), such reconceptualising of how safety is constructed on the construction frontline will provide the industry with additional means by which safety can be understood and managed.

References

Braun, V., Clarke, V.: One size fits all? What counts as quality practice in (reflexive) thematic analysis? Qual. Res. Psychol. (2020). https://doi.org/10.1080/14780887.2020.1769238

Davies, A. (2017), Projects: A Very Short Introduction, Oxford University Press, Oxford, UK.

Furniss, D. *et al.* (2019) 'Exploring structure, agency and performance variability in Everyday Safety: An ethnographic study of practices around infusion devices using distributed cognition', *Safety Science*, 118, pp. 687–701. doi:10.1016/j.ssci.2019.06.006.

Han, S., Saba, F., Lee, S., Mohamed, Y., & Peña-Mora, F. (2014). Toward an understanding of the impact of production pressure on safety performance in construction operations. Accident; Analysis and Prevention, 68, 106–116. https:// doi.org/10.1016/j.aap.2013.10.007.

Haslam, R.A. *et al.* (2005) 'Contributing factors in construction accidents', *Applied Ergonomics*, 36(4), pp. 401–415. doi:10.1016/j.apergo.2004.12.002.

Hollnagel, E. (2018), *Safety-II in Practice*, Routledge, Abingdon, UK.

HSE (2024), 'Statistics - Work-related fatal injuries in Great Britain', https://www.hse.gov.uk/statistics/fatals.htm. Accessed: 23rd November 2024.

McDonald, N. (2006), Organisational Resilience and Industrial Risk, in E. Hollnagel, D. D. Woods and N. Levenson, eds, 'Resilience Engineering: Concepts and Precepts', 1 edn, Ashgate Publishing Limited, Aldershot, UK, chapter 11, pp. 155–179.

Rae, A. *et al.* (2020) 'A manifesto for reality-based Safety Science', *Safety Science*, 126, p. 104654. doi:10.1016/j.ssci.2020.104654.

Reiman, T., Rollenhagen, C., Pietikäinen, E., & Heikkilä, J. (2015). Principles of adaptive management in complex safety–critical organizations. *Safety Science*, *71*, 80–92. doi:10.1016/j.ssci.2014.07.021

Tversky, A. and Kahneman, D., 1974. Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, *185*(4157), pp.1124-1131.

Woods, D. D. and Hollnagel, E. (2006), Prologue: Resilience Engineering Concepts, in E. Hollnagel, D. D. Woods and N. Levenson, eds, 'Resilience Engineering: Concepts and Precepts', 1 edn, Ashgate Publishing Limited, Aldershot, UK, chapter Prologue, pp. 1–6.

Adopting Passive Exoskeletons: Worker Perspectives and Impact on Work Productivity and Quality

Amin Yazdani & Marcus Yung

Canadian Institute for Safety, Wellness, and Performance, Conestoga College, Canada

SUMMARY

Exoskeletons help augment worker strength and may reduce the effects of physically demanding work; however, health benefits alone may not necessarily facilitate its adoption by organizations. Through a survey study with 40 construction workers who used an exoskeleton over multiple shifts, workers believed exoskeletons were usable and beneficial for prolonged overhead work. A subsequent systematic review suggests that quality and productivity impacts of exoskeleton use are dependent on task characteristics.

KEYWORDS

Economic impact, usability, inclusive design

Introduction

Passive exoskeletons are wearable devices that are designed to augment a worker's strength. In laboratory studies, passive exoskeletons may increase muscular endurance. Our recent multi-day field studies have shown the potential of passive exoskeletons to reduce bodily exertion and discomfort in construction work. Therefore, exoskeletons are perceived as a promising solution to reduce the risk of musculoskeletal disorders (MSD) that are highly prevalent and costly in Canada.

Despite emerging evidence that exoskeletons may reduce the effects of high physical demands, the health benefits alone may not necessarily facilitate their adoption by an organization. An exoskeleton's practical use and application for work tasks as well as worker perceptions on its comfort and usability are important factors for adoption. Additionally, improvements to quality, productivity, and reduced costs are powerful business agendas that would likely receive more resources and attention. This study examined the perceptions of construction workers who have used an exoskeleton for multiple work shifts to identify tasks that may benefit from exoskeletons and identify design improvements to facilitate their adoption. We also report on a systematic review of scientific literature to synthesize the current state of knowledge on the impact of exoskeletons on work productivity, quality, and their economic implications.

Method

We recruited 40 residential, ICI (Industrial, Commercial, Institutional), and modular construction workers to participate in a survey study. Participants were surveyed on their job sites after using a passive upper extremity exoskeleton (Hilti Exo-01) to perform their jobs for multiple work shifts. We elicited participants' perceptions on the usability of exoskeletons, the features of an exoskeleton most beneficial for overhead work, the types of construction tasks suitable for upper extremity exoskeletons, and design recommendations to improve exoskeleton adoption.

Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, we identified, reviewed and extracted data from journal articles obtained from six large databases. We included research articles that focused on exoskeletons worn during occupational tasks and that had evaluated its effects on quality and productivity, or its economic impacts.

Preliminary Findings

Eighty percent of participating construction workers believed the passive upper extremity exoskeleton was easy to use. Participants preferred the exoskeleton for the reduction in muscular effort, their looks, and feeling safe, but preferred their usual method for manoeuvrability, thermal comfort, and feel/fit (Figure 1). Workers believed exoskeletons were helpful for prolonged, static overhead tasks in open spaces, such as installing upper tracks and light fixtures, framing and drywalling bulkheads, and taping and mudding ceilings. On the other hand, exoskeletons were viewed as cumbersome in tight spaces or below shoulder height. Participants recommended better design compatibility with tools and PPE, close-fitting design for confined spaces, breathable materials, debris prevention, and better fit to accommodate different body sizes and dimensions.

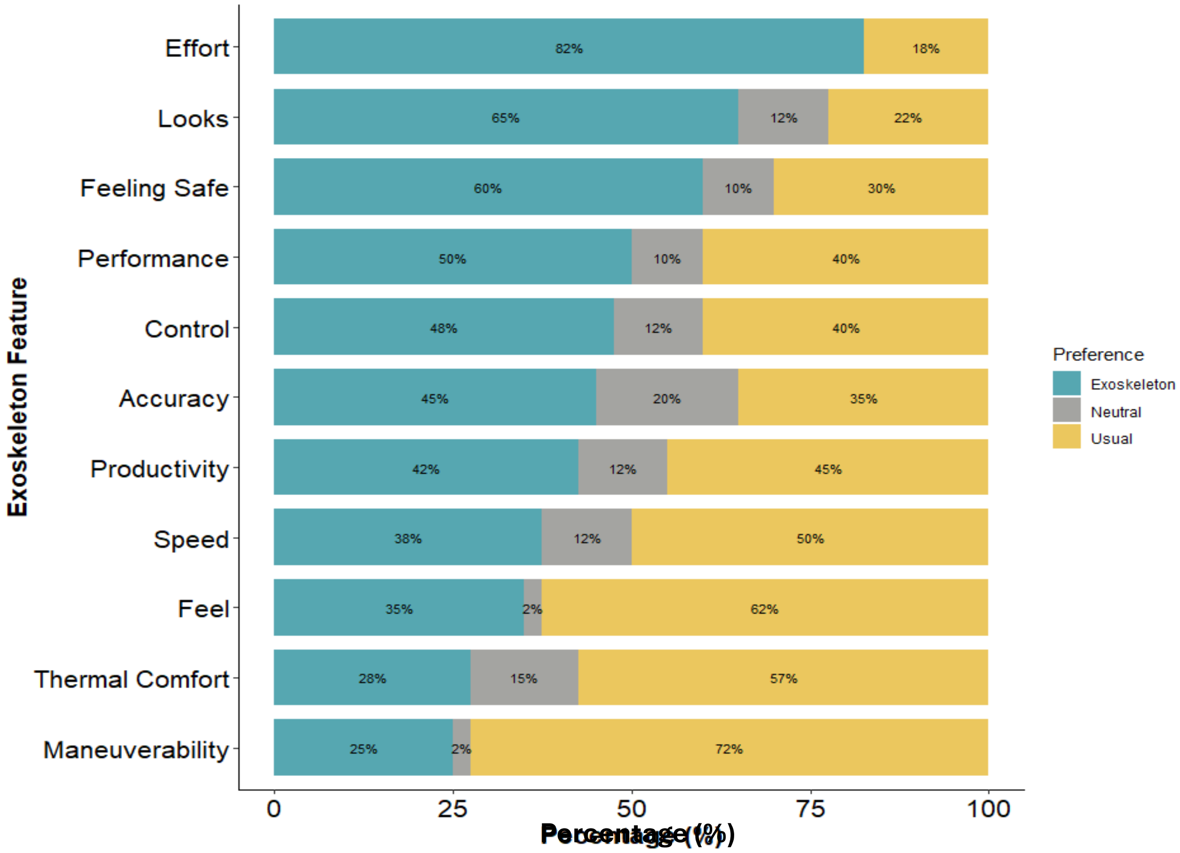


Figure 1: Preferences of using an exoskeleton for overhead work compared to usual methods.

Fifteen articles examined the effects of exoskeleton on work productivity and quality. Exoskeleton use positively affected productivity in certain tasks such as shovelling, welding, and electrostatic painting. However, there were mixed productivity findings during simulated manual material handling, drilling, and manual assembly tasks. Welders and electrostatic painters experienced quality improvements in their task, while the impact of exoskeleton on committing errors during repetitive drilling tasks was dependent on work height. No articles assessed the economic impact.

Key Takeaways

- Exoskeletons were perceived to be usable by almost 80% of our 40 sampled construction workers who used a passive upper extremity exoskeleton for multiple work shifts. However, there remains issues with manoeuvrability, thermal discomfort, and fit/feel.
- Upper extremity exoskeletons may be most beneficial for tasks requiring prolonged, static overhead work in open spaces.
- The current state of the literature suggests that quality and productivity impacts of exoskeleton use are dependent on task characteristics.
- There remains a need for cost-benefit analysis and studies on return on investment to justify the benefits of exoskeleton adoption for the organization and its workers.
- Future studies should continue to consider evaluating exoskeletons with actual workers and identify possible effects of sex and gender.

Fatigue Risk Management within the UK Construction Industry

Shelley Stiles

Gateway Consultants (HSW) Ltd

SUMMARY

Fatigue is reported an important contributory factor in safety incidents across UK, including within the construction industry. The aim of this study is to gain a better understanding of existing organisational arrangements for the management of fatigue risk within the UK construction sector and compare maturity of the approach with other industries.

KEYWORDS

Fatigue, construction, organisational maturity

Fatigue and UK Construction

The UK Construction industry has one of the highest rates of poor psychosocial health, including fatigue. Fatigue is considered to be a decline in mental and/or physical performance that results from prolonged exertion, sleep loss and/or disruption of a person's 'internal clock'. Worker fatigue can result in a lack of attention, slower reactions, reduced co-ordination, decreased awareness, underestimation of risk, memory lapses or absent-mindedness and a reduced ability to process information.

Fatigue can arise as a result of excessive working time or poorly designed shift patterns. It is also related to workload, since workers are more easily fatigued if their work is complex or monotonous. The characteristics of the construction industry predispose conditions for fatigue associated with a peripatetic workforce, skills shortage, national contracts, and a high proportion of self-employed workers. An industry study by Considerate Constructors Scheme established that long working hours (86% more than 40 hours per week), extensive routine travel (44% travelling to work 2 to 3 hours each day), high workload demands both cognitive and physical are the norm across the sector.

Fatigue management in other industries

There are other industries such as aviation, rail, oil and petrochemical, healthcare, transport and logistics, which require a fatigue risk management system (FRMS) as part of the normal licence to operate within those sectors. Key components of a FRMS include policy, objectives, risk management, assurance and promotion of fatigue. Often this more stringent approach has followed major disasters where fatigue has been identified as a causal or contributory fatigue such as the Challenger Space Shuttle disaster in 1986, or Clapham rail disaster in 1988, where the ensuing investigations identified significant changes to each industry's approach to fatigue. There has not been a similar disaster within the construction sector where fatigue has been identified as a causal factor, and therefore the licences to operate within the sector are less stringent.

The aim of this study is to gain a better understanding of organisational arrangements for the management of fatigue risk within the construction sector, in comparison with other sectors.

Study Approach

The study was undertaken in two parts, firstly an online survey of organisational arrangements, and secondly comparison with FRMS information from other industries using industry legislation, standards, and guidance. The industries included were aviation, rail, oil and petrochemical, healthcare, transport and logistics. This paper provides details on part one only.

Part one of the study involved an online survey of 19 organisations from within the construction sector to identify their approach to fatigue risk management including fatigue policy, fatigue training. There were organisations who worked within the civils, residential and commercial building and infrastructure sectors of construction and included Principal Designer, Designer, Principal Contractor and Contractor roles on projects. The organisations all employed more than 10 people, with a turnover more than £2 million per year.

Study Findings

Of the 19 organisations, many have fatigue related policies in place which are applied across the organisation including construction projects and the subcontract workforce.

- 64% have a fatigue risk management policy, with 41% having a night worker policy
- 82% have a driver fatigue management policy
- 50% had a travel and/or commuting policy (most organisations allowed a commute time of 30 minutes to 2 hours each way for each shift)

In terms of typical shift duration, most organisations (84%) rostered daytime shifts of circa 10 hours, with 31% rostering shifts of more than 12 hours. 42% also rostered night shifts (between 18.00 and 06.00). The majority also defined the minimum rest period between shifts as 12 hours, and one day off each week. Commuting/travelling was included within roster planning. However, more than half of organisations did not undertake specific fatigue risk assessments as part of their roster planning.

68% of organisations provide fatigue management training to their teams, although those working solely in building (both residential and commercial) provide no specific fatigue awareness training. Organisations employing less than 250 people with an annual turnover less than £25million all had training in place, where not all of the larger organisations had this in place. Beyond having fatigue-related policies and training in place, most organisations did not involve their workforce in designing and reviewing rosters, have established fatigue performance indicators, or make provisions to aid individual wellbeing (exercise, nutrition, quality of rest).

Conclusions

Conclusions are based on part one of this study with 19 participant organisations, recognising that such a limited number may not be representative of the whole industry. This study has found that the majority of organisations working in the infrastructure and civil engineering sectors of the construction industry have fatigue related policies in place, specifying hours of work (including travel and commuting) and they provide fatigue training to their teams. Small to medium-sized organisations also have more fatigue-related arrangements in place, probably due to their direct employment of those frontline construction workers who are exposed to the greater fatigue risk from both physical and cognitive demand. FRMS are less established within the building (both residential and commercial) sectors.

Reference

Considerate Constructors Scheme https://ccsbestpractice.org.uk/spotlight-on/spotlight-on-worker-fatigue/#Introduction

WRMSD risks within the construction industry

Andrew D.J. Pinder

Health and Safety Executive, UK

SUMMARY

The study sought to review in a systematic way the evidence comparing the rates of work-related musculoskeletal disorders in different construction occupations. The available data indicate that some occupations are significantly worse than others. The samples underlying the data are not sufficiently large to produce reliable estimates of prevalences in all but the largest occupations.

KEYWORDS

Construction, Work-related musculoskeletal disorders, Health and safety

Background

HSE tracks work-related ill-health through the annual Labour Force Survey (LFS). It estimates that 543 000 workers suffered from, and 7.8 million working days were lost, due to new or long-standing work-related musculoskeletal disorders (WRMSDs) in 2023/24 . WRMSDs such as low back pain are often associated with manual handling (such as lifting, carrying, pushing or pulling) of loads.

The LFS shows that the construction industry has a much higher rate of WRMSDs (approximately 1.7 times for the period from 2021/22 to 2023/24) than the average across all industries. Some tasks within construction create higher risks of WRMSD than others.

Aim

The specific aim of the project was to search for, collate and analyse existing information, particularly published scientific studies linking particular construction tasks and trades with increased risk of WRMSD risks in construction in order to rank tasks and activities by risk of WRMSDs.

Approach

The planned approach of the project was to carry out a systematic review of scientific literature on WRMSDs related to the construction sector, concentrating on the higher quality epidemiological studies. A search of databases of scientific literature for a wide range of terms related to "musculoskeletal disorders" AND "the construction industry" identified over 2000 publications. There were no geographical, language or date restrictions, except that the search terms were in English. Figure 1 is a PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses (Page *et al.*, 2021)) flow chart for the study, showing how relevant papers were identified.

Retrieved records were managed in EndNote (version X9, and later version 20), which automatically downloads electronic copies of Open Access papers. The three members of the study team worked independently to classify the 1708 papers remaining after removal of duplicates and irrelevant papers. Studies reporting epidemiological data on individual construction trades or tasks were included and classified by the type of epidemiological study. Studies of the construction sector as a whole without any breakdown were excluded as it is already known that construction is high risk compared to other sectors.

The study team then considered which articles should be included in the planned review by starting with the systematic reviews (the highest quality type) and working down the study hierarchy (Greenhalgh, 1997) until sufficient quality information was found that could be extracted that addressed the study question. Differences between the classifications of the three team members were resolved in team meetings. At this stage, titles, abstracts and, where already available, the full papers, were taken into consideration. Table 1 shows the numbers of papers included at each stage for the six main epidemiological study types identified. Each paper that was chosen for full-text review was allocated to two of the three team members who independently extracted relevant details into Excel spreadsheets. The team then met to compare findings and agree how to proceed.

Table 1: Results of the screening and selection of studies

Study type	Papers identified by the three reviewers from the titles and abstracts	Papers chosen for full-text review	Papers chosen for data extraction
Systematic reviews	9	5	0
Intervention studies	150	8	0
Prospective cohort studies	70	41	0
Registry/retrospective studies	63	40	23
Case-control studies	16	NA	NA
Cross-sectional studies	190	NA	NA

Nine papers were identified as reporting systematic reviews, of which five were chosen as suitable for full-text review. None were chosen for data extraction and, given the limited number of reviews found, their heterogeneity, their limited conclusions and the lack of data on individual trades and tasks, the team agreed that it was necessary to proceed to reviewing the intervention studies identified in the search.

The team had identified 150 papers as possible intervention studies. Of these, only eight were chosen for full review, but this found that only three had potentially useful information. Because the number was so small, the team agreed to proceed to reviewing prospective cohort studies. Of the 70 potential prospective cohort papers, 41 were chosen for full-text review. Of these, eleven, reporting six different studies, were identified as suitable for data extraction. However, the review team agreed that the studies were quite varied, and viewed overall, provided insufficient data to allow conclusions to be drawn about individual construction trades and tasks.

Through a process of sorting and evaluation, it was decided to concentrate on studies that examined data collected through injury and compensation claim and medical surveillance databases over periods of years (registry/retrospective longitudinal studies). Typically, these databases are available in countries that use Workers' Compensation (WC) insurance systems to pay for medical care / lost time when a worker is injured at work. They can include very large numbers of workers, but are limited when significant numbers of workers opt for other types of health cover.

Sixty-three studies were initially identified as registry/retrospective studies, and 40 identified as suitable for retrieval of the full-text. One could not be retrieved; one was found to be only a conference abstract and two did not report data specific to construction trades. Full text review of the remaining 36 reduced the number under consideration to 23. Of the 13 rejected, four had too limited data; three did not report WRMSD data separately to other injury data; two reported data also reported in related papers; one did not report data specific to construction trades; one did not report injury rate data; one reported data from a single trade and in one the data were very old in an area (hand tools) that has seen significant design changes since the paper was published.

Relevant data from the 23 papers were copied into Excel workbooks and examined in detail. After extraction of the data from the 23 papers, four were excluded, one due to duplicate publication, one due to not reporting separate WRMSD data and two due to not reporting incidence data. This meant that usable data were obtained from 19 papers. Four papers (Lipscomb *et al.*, 1997, Lipscomb *et al.*, 2008a, b, Lipscomb *et al.*, 2015) relating to WC insurance claims by unionised carpenters in Washington State in the USA. Five related to general WC claims made in Washington State (Silverstein *et al.*, 2002, Bonauto *et al.*, 2006, Schoonover *et al.*, 2010, Spector *et al.*, 2011, Anderson *et al.*, 2013). Two papers reported data from the US Bureau of Labor Statistics (BLS) annual Survey of Occupational Injuries and Illnesses (SOII) (Wang *et al.*, 2017, Dong *et al.*, 2019). A pair of papers (Kontio *et al.*, 2018, Solovieva *et al.*, 2018) reported rates of, and risk factors for, knee and hip osteoarthritis in Finland. Six papers were unrelated to the other included studies. Duguay *et al.* (2001) reported WC claim data from Quebec. Stocks *et al.* (2011) analysed cases of work-related ill-health reported to HSE. Andersen *et al.* (2012) analysed cases of hip and knee osteoarthritis in Denmark. Wahlström *et al.* (2012) discussed Swedish cases of lumbar disc disease that resulted in hospitalisation. Memarian and Mitropoulos (2013) analysed recordable incidents that occurred in a large construction company specialising in brick and blockwork. The final paper, Dale *et al.* (2015), compared the WRMSD medical claims for floor layers in Missouri with those for the general working population.

Examining the papers by Wang *et al.* (2017) and Dong *et al.* (2019) led to a decision to include more comprehensive US SOII data and therefore to draw on the similar GB LFS data. Injury reports/claims are classified by both the economic sector of the employer, eg using Standard Industrial Classification (SIC) codes, and by occupational groups, eg using Standard Occupational Classification (SOC) codes. This allows injuries to be considered by both the type of business employing the injured person and by the kind of work the injured person performs.

Separate LFS data files for illnesses and injuries analysed by both UK SIC code and UK SOC code were downloaded from the HSE statistics microsite, https://www.hse.gov.uk/statistics/index.htm. The SIC data are broken down to the lowest level of industry group. The SOC data are broken down to the smallest occupational group. For many of the lower level groups, the sample sizes were too small for HSE to be able to provide reliable estimates. The available prevalence / incidence data and prevalence ratio / incidence ratio data and associated 95% Confidence Intervals were extracted and the numbers of FTE workers in each occupational class estimated from them.

The UK SIC code data used were annual arithmetic means over the three year period from April 2016 to March 2019. The UK SOC code data used were the annual means over the period from April 2017 to March 2020. These two slightly different periods were selected as being recent but not significantly affected by the Coronavirus pandemic that started in early 2020.

The SOII data from the USA were downloaded from www.bls.gov for the ten year period from 2011 to 2020. Injury and illness rates were analysed by 35 North American Industry Classification System (NAICS) codes within the construction sector code of 23XXXX. Of these, 22 were ‘industries’ (codes 23XXX0), ten were ‘industry groups’ (codes (23XX00) and three were ‘subsectors’ (codes 23X000). The same data were also analysed by 64 construction occupations (US 2010 SOC code) in SOC codes 47-0xxx to 47-4xxx plus construction-related occupations in other top-level occupational groups, such as construction managers (SOC code 11-902x).

Rates from all sources were standardised as rates per 100 Full-Time Equivalent workers / per 200,000 hours worked, using a notional full-time rate of 2000 hours worked per year.

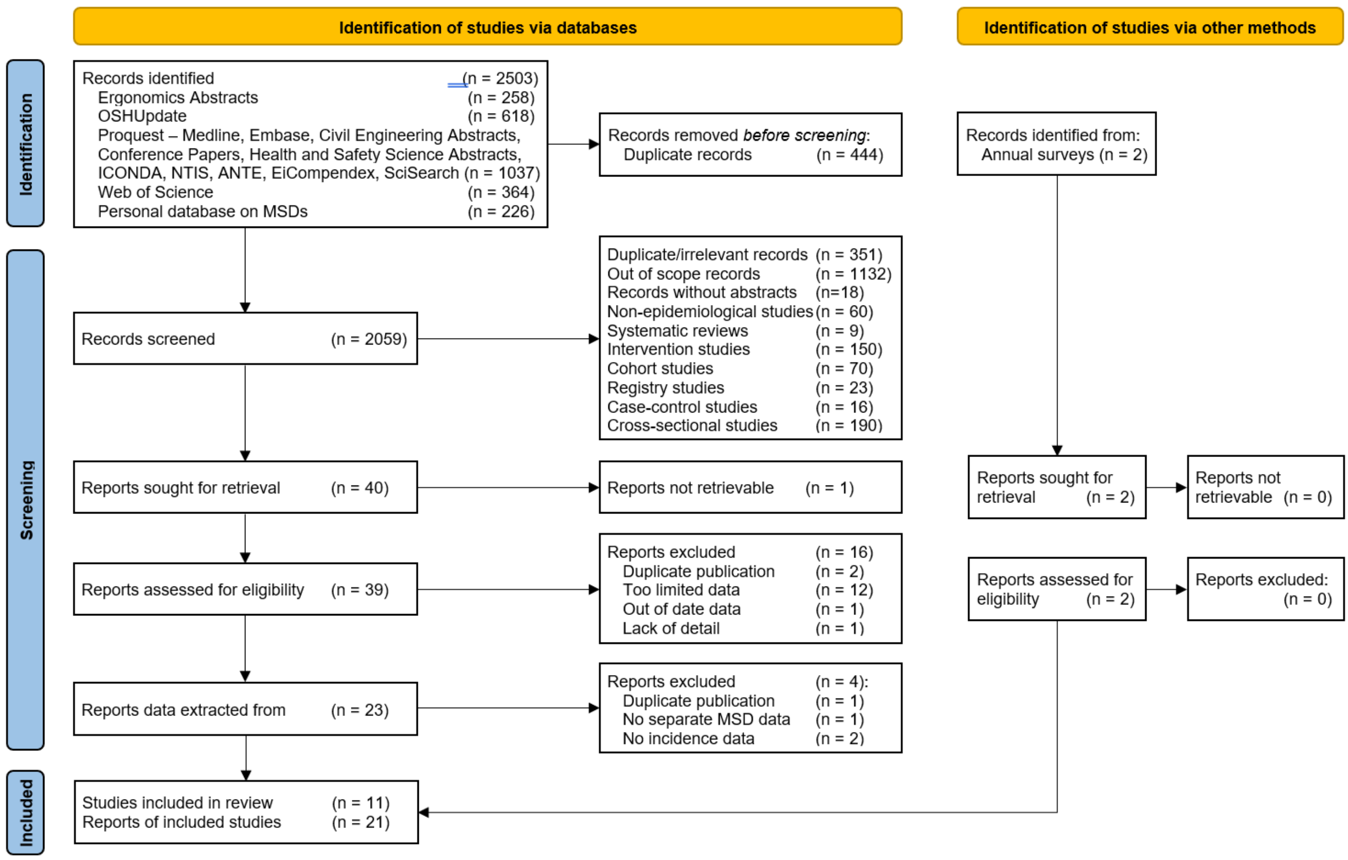


Figure 1: PRISMA 2020 flow diagram for the study (Page *et al.* (2021), licensed under CC BY 4.0)

Figure 1 is a flow diagram summarising the results of the search and selection process, from the number of records identified in the search to the number of studies finally included in the review. It uses the format provided by Page *et al.* (2021) as part of the recommendations of the *Preferred Reporting Items for Systematic reviews and Meta-Analyses* (PRISMA) statement.

Findings

Only limited findings from the analysis are reported here. They form part of the findings discussed in Pinder *et al.* (In press).

The LFS data indicate that the prevalence of WRMSDs in the UK construction sector is 1.8 times (= 2.11/1.16) the prevalence averaged across all industries. The limitation of the LFS sample size meant that only seven of 33 unique UK SIC 2007 construction industry codes had sufficient data for reliable estimates to be provided in relation to WRMSDs. Nor did the sample size allow detailed analysis by affected body parts. Only two SIC codes were at the lowest level of ‘industry classes’: UK SIC 2007 code F41.20 ‘Construction of residential and non-residential buildings’ (1.67 reports per 200 000 hours worked) and F43.32 ‘Joinery installation’ (3.68 reports per 200 000 hours worked). F43.3 ‘Building completion and finishing’ has twice (= 4.17/2.11) the prevalence of WRMSDs across the whole sector. F43.3 ‘Building completion and finishing’ has 2.2 times the rate (= 4.17/1.93) of code F43.2 ‘Electrical, plumbing and other construction installation activities’.

Of the 62 UK SOC 2010 occupational groups relevant to construction, only 19 have sufficient LFS data for any estimates to be provided in relation to WRMSDs. The three (of 33) unit groups with sufficient data all had high WRMSD rates and were all within sub-major group 53 ‘Skilled construction and building trades’ / minor group 531 ‘Construction and building trades’: 5314

‘Plumbers and heating and ventilating engineers’; 5315 ‘Carpenters and joiners’; and 5319 ‘Construction and building trades n.e.c.’ (n.e.c. = not elsewhere classified).

US SOII data showed that averaged from 2011 to 2020, the four worst construction industries for WRMSD reports were NAICS codes 23835 ‘Finish carpentry contractors’ (incidence rate (IR) = 0.563 absences per 200 000 hours), 23833 ‘Flooring contractors’ (IR = 0.561), 23813 ‘Framing contractors’ (IR = 0.551) and 23817 ‘Siding contractors’ (IR = 0.539). The first two of these map onto UK SIC 2007 codes F43.32 ‘Joinery installation’ and F43.33 ‘Floor and wall covering’ respectively. The other two both map onto F43.99 ‘Other specialised construction activities n.e.c.’. The six worst construction occupations over that period were US 2010 SOC codes 47-4023 ‘Floor sanders and finishers’ (IR = 3.581 absences per 200 000 hours),47-2042 ‘Floor layers, except carpet, wood and hard tiles’ (IR = 1.119), 47-2142 ‘Paperhangers’ (IR = 1.058), 47-2041 ‘Carpet installers’ (IR = 0.966), 49-2022 ‘Telecommunications equipment installers and repairers, except line installers’ (IR = 0.866), and 49-9021, ‘Heating, air conditioning, and refrigeration mechanics and installers’ (IR = 0.849). Codes 47-4021, 47-2042 and 47-2043 map to UK SOC 2010 code 5322 ‘Floorers and wall tilers’, 47-2142 maps to 5323 ‘Painters and decorators’, 49-2022 maps to 5242, ‘Telecommunications engineers’ and 49-9021 maps to 5225 ‘Air-conditioning and refrigeration engineers’

Of the 66 construction occupational groups in the US 2010 SOC, 54 (82%) had data from at least seven years in the period from 2011 to 2020, and 11 of these had statistically significant trends in incidence rates, all downward. Of these, eight were ‘detailed occupations’, the lowest level of classification. Correlations between year and incidence rate ranged between -0.70 and -0.93, so the amount of variability associated with the time trend ranged between 49% and 86%, indicating that these were consistent decreases in incidence rates. Such decreases over a 10-year period would suggest that longer-term trends are reducing WRMSD risks in these occupations.

The studies of unionised carpenters in Washington State (Lipscomb *et al.*, 1997, Lipscomb *et al.*, 2008a, b, Lipscomb *et al.*, 2015) indicated the highest risk was to workers performing drywall work. Residential and light commercial carpentry work was also at greater risk than other carpentry work. The data of Silverstein *et al.* (2002) show that certain construction activities were high risk for multiple body parts, suggesting that they were creating excessive demands on the whole body. These were Washington Industrial risk Classification (WIC) risk classes 0515 ‘Wallboard installation’, 0518 ‘Building construction, NOC’ and 0510 ‘Wood frame building construction’. Reinforcing steel installation was the most hazardous risk class for carpal tunnel syndrome.

There is evidence (Andersen *et al.*, 2012) that individual construction trades have rates of musculoskeletal problems higher than for construction workers as a whole, for both work absences and increased rates of surgery for osteoarthritis (such as hip replacements) among floor layers, brick layers and pavers . Similarly, Solovieva *et al.* (2018) found skilled construction workers, such as electricians and plumbers, had higher rates of disability due to hip osteoarthritis than unskilled transport, construction and manufacturing workers. They also found that combining kneeling and squatting and heavy physical work increased the risk. This suggests that prevention strategies could focus on the risks of heavy physical work that occurs in awkward postures.

Conclusions

A large number of registry/retrospective studies reporting longitudinal WRMSD data for construction industry groups and construction trades were identified as useful for HSE’s purpose of trying to identify construction trades and occupations at increased risk of WRMSDs. They largely came from North America and Europe. Most studies from the USA used data from insurance-based WC systems, so reported Claims Incidence Rates due to WRMSDs, while some European studies report hospitalisation rates for specific problems such as ‘lumbar disc disease’ or hip or knee osteoarthritis. The WC data exclude uninsured workers and are likely to exclude low-severity incidents that do not lead to claims. Annual survey data were also obtained from both the UK LFS and the US SOII.

Rates of injury are unrelated to the size of the construction industry group/trade concerned, meaning that both small and large groups of workers could have either low or high injury rates. Days absent from work due to WRMSDs were also unrelated to the size of the construction industry group/trade and its injury rate.

Cumulative / repetitive injury WRMSDs were shown to be much more common and more disabling than 'acute onset' or overexertion injuries. This suggests that prevention efforts should be focussed on reducing long-term exposures to WRMSD risk factors, such as lifting heavy weights, while not ignoring risk factors for acute injuries, such as extreme loads.

Methodological issues

Some of the studies from which data were extracted had limitations in their methodologies that made it very difficult to draw generalisable conclusions from them. For instance, Stocks *et al.* (2011) drew data from clinical reports of work-related illness via a voluntary reporting system that appears not to have been able to provide a large representative sample of the construction workforce.

A number of studies used construction sector-wide, or population databases to identify clinical outcomes and therefore have the advantage of comprehensive coverage of the at-risk population. However they often used very restrictive case definitions, such as hospitalisation (eg, Wahlström *et al.*, 2012). This will result in lower incidence rates than in employer surveys such as the US SOII (Dong *et al.*, 2019, Wang *et al.*, 2017) that include cases of WRMSDs with less severe outcomes, such as work loss. However, such surveys can suffer from differential reporting by employers.

The SOII data were quite variable from year to year, meaning that data averaged over multiple years give more reliable comparisons than the single year reported by Dong *et al.* (2019). The SOII data for days absent from work were even more variable, but their data for numbers of cases reported and hence injury rates were much more stable. This led to the conclusion that injury rates are the preferable measure when comparing industries and occupations.

Key takeaways

Classification systems evolve over time to reflect changes in the economy and in individual occupations, so it can be difficult to make comparisons over extended periods.

The WRMSD data available for the UK construction industry are limited in detail, so make detailed analysis at the level of occupation or employer type very challenging. In particular, the LFS sample size is too small for HSE to make reliable estimates of WRMSD prevalences in all but the largest construction industry occupations and industrial groupings. However, there is sufficient data to indicate that HSE should consider focussing on the building completion and finishing group of industries (UK SIC group F43.3) as a whole, and on the UK SOC classes 5314 'Plumbers and heating and ventilating engineers', 5315 'Carpenters and joiners' and 5319 'Construction and building trades n.e.c.' in particular.

US data show a clear trend in the reduction of WRMSD reporting incidence rates over many years and across their economy, not just in their construction sector. This suggests that wider societal trends in materials handling equipment, employment and/or health may have had positive effects on musculoskeletal health among workers in the USA. The rates of reduction varied between occupations, suggesting that specific factors have reduced risk differentially. There is insufficient evidence available to show if such trends have occurred in the UK.

Ranking data from multiple industries or occupations by incidence rate shows that though the overall range is large, the steps between neighbouring groups are often small, and there are no distinct clusters. This suggests that focusing health and safety interventions on the highest risk occupations as the most likely way to reduce risk may not show clear-cut results because the differences between different occupations may be small.

Acknowledgements

This paper and the work it describes were funded by the Health and Safety Executive (HSE). Its contents, including any opinions and/or conclusions expressed, are those of the author alone and do not necessarily reflect HSE policy.

References

Andersen, S., Thygesen, L.C., Davidsen, M., *et al.* (2012). Cumulative years in occupation and the risk of hip or knee osteoarthritis in men and women: a register-based follow-up study. *Occup Environ Med,* **69**, (5)**,** 325-330.

Anderson, N.J., Bonauto, D.K. and Adams, D. (2013). *Prioritizing industries for occupational injury and illness prevention and research, Washington State workers' compensation claims data, 2002-2010.* (Washington State Department of Labor and Industries), Technical Report Number 64-1-2013, 58 page(s). https://lni.wa.gov/safety-health/safety-research/files/2013/bd_3F.pdf.

Bonauto, D., Silverstein, B., Adams, D., *et al.* (2006). Prioritizing Industries for Occupational Injury and Illness Prevention and Research, Washington State Workers' Compensation Claims, 1999–2003. *J Occup Environ Med,* **48**, (8)**,** 840-851.

Dale, A.M., Ryan, D., Welch, L., *et al.* (2015). Comparison of musculoskeletal disorder health claims between construction floor layers and a general working population. *Occup Environ Med,* **72**, (1)**,** 15-20.

Dong, X.S., Betit, E., Dale, A.M., *et al.* (2019). *Trends of musculoskeletal disorders and interventions in the construction industry.* (CPWR-The Center for Construction Research and Training), 20 page(s). https://www.cpwr.com/wp-content/uploads/2020/06/Quarter3-QDR-2019.pdf

Duguay, P., Cloutier, E., Levy, M., *et al.* (2001). Statistical profile of compensated back injuries in the Quebec construction industry. [Original French title = Profil statistique des affections vertébrales avec indemnités dans l'industrie de la construction au Québec.] *Trav Hum,* **64**, (4)**,** 321-342.

Greenhalgh, T. (1997). How to read a paper. Getting your bearings (deciding what the paper is about). *BMJ,* **315**, (7102)**,** 243-246.

Kontio, T., Viikari-Juntura, E. and Solovieva, S. (2018). To what extent do education and physical work load factors explain occupational differences in disability retirement due to knee OA? A nationwide register-based study in Finland. *BMJ Open,* **8**, (11)**,** e023057.

Lipscomb, H.J., Cameron, W. and Silverstein, B. (2008a). Back injuries among union carpenters in Washington State, 1989-2003. *Am J Ind Med,* **51**, (6)**,** 463-474.

Lipscomb, H.J., Cameron, W. and Silverstein, B. (2008b). Incident and recurrent back injuries among union carpenters. *Occup Environ Med,* **65**, (12)**,** 827-834.

Lipscomb, H.J., Dement, J.M., Loomis, D.P., *et al.* (1997). Surveillance of work-related musculoskeletal injuries among union carpenters. *Am J Ind Med,* **32**, (6)**,** 629-640.

Lipscomb, H.J., Schoenfisch, A.L., Cameron, W., *et al.* (2015). Workers' Compensation Claims for Musculoskeletal Disorders and Injuries of the Upper Extremity and Knee Among Union Carpenters in Washington State, 1989-2008. *Am J Ind Med,* **58**, (4)**,** 428-436.

Memarian, B. and Mitropoulos, P. (2013). Accidents in masonry construction: The contribution of production activities to accidents, and the effect on different worker groups. *Saf Sci,* **59,** 179-186.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., *et al.* (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ,* **372,** n71.

Pinder, A.D.J., Wood, L. and Yeomans, E. (In press). *Work-related musculoskeletal disorders within the construction sector – a systematic review.* (HSE Books), RR1220. https://www.hse.gov.uk/research/rrhtm/index.htm.

Schoonover, T., Bonauto, D., Silverstein, B., *et al.* (2010). Prioritizing prevention opportunities in the Washington State construction industry, 2003-2007. *J Safety Res,* **41**, (3)**,** 197-202.

Silverstein, B., Viikari-Juntura, E. and Kalat, J. (2002). Use of a prevention index to identify industries at high risk for work-related musculoskeletal disorders of the neck, back, and upper extremity in Washington State, 1990-1998. *Am J Ind Med,* **41**, (3)**,** 149-169.

Solovieva, S., Kontio, T. and Viikari-Juntura, E. (2018). Occupation, physical workload factors, and disability retirement as a result of hip osteoarthritis in Finland, 2005-2013. *J Rheumatol,* **45**, (4)**,** 555-562.

Spector, J.T., Adams, D. and Silverstein, B. (2011). Burden of Work-Related Knee Disorders in Washington State, 1999 to 2007. *J Occup Environ Med,* **53**, (5)**,** 537-547.

Stocks, S.J., Turner, S., McNamee, R., *et al.* (2011). Occupation and work-related ill-health in UK construction workers. *Occup Med,* **61**, (6)**,** 407-415.

Wahlström, J., Burström, L., Nilsson, T., *et al.* (2012). Risk factors for hospitalization due to lumbar disc disease. *Spine,* **37**, (15)**,** 1334-1339.

Wang, X.W., Dong, X.W.S., Choi, S.D., *et al.* (2017). Work-related musculoskeletal disorders among construction workers in the United States from 1992 to 2014. *Occup Environ Med,* **74**, (5)**,** 374-380.

Design And Evaluation of An Ergonomics Risk Assessment Report Using Sensor-Based Data

Liyun Yang[1], Andreas Hallberg[2], Ida-Märta Rhen[1,2], Farhad Abtahi[3,4], Jörgen Eklund[1] & Mikael Forsman[1,2,3]

[1] Institute of Environmental Medicine, Karolinska Institutet, Sweden; [2] Centre for Occupational and Environmental Medicine, Stockholm County Council, Sweden; [3] Division of Ergonomics, KTH Royal Institute of Technology, Stockholm, Sweden; [4] Department of Clinical Science, Intervention and Technology, Karolinska Institutet, Sweden

SUMMARY

The objective of this paper is to present the design and evaluation results of three digital ergonomics risk assessment reports based on sensor-measured data of upper arms, trunk, head and the dominant wrist. An iterative design process involving industrial user groups is used. Fifteen industrial users (five women) answered a digital survey presenting three designs of the ergonomics risk assessment report. The Interface Usability Instrument (INUIT) and open questions on suggestions were included in the survey. The total INUIT scores of the report design A, B, and C were 71, 64, and 61 (normalised to 100 point). Design A showed best ratings regarding confusion and distraction of the three designs. All three designs were considered supportive in communicating the ergonomic risks. In conclusion, the designed reports can be used to support end users in interpreting and communicating ergonomics risks, comparing and improving design of workstations or aids, or evaluating work techniques. The methodology ensures active collaboration between researchers and industrial stakeholders to understand user needs and facilitate the implementation of such digital reports within the industry. In the next stage of the project, the report will be further improved, implemented in the online digital platform, and tested by users in real work scenarios.

KEYWORDS

Ergonomics risk assessment, wearable system, report design

Introduction

There is rapid development in smart wearables and the advancement of applying such technology for ergonomics (Lim and D'Souza, 2020; Lind et al., 2023). Technical measurements using sensors can provide data of higher accuracy for ergonomics risk assessment compared to self-reported and observational data (Hansson et al., 2001; Takala et al., 2010). The latest wearable systems also offer the possibility of longer and continuous risk assessments, which are convenient to measure and analyse, some of which provide results that are analysed directly after the measurement. The DiPMaS (Health, Safety, and Quality in a Digital Platform for Manufacturing Staff) project is a consortium between academia and industrial organisations with an aim to develop a digital platform using smart workwear systems with integrated sensors to support manufacturing staff to improve health, safety and quality. The current system uses inertial measurement unit sensors, a mobile application, and a cloud platform to analyse data and visualise risk assessment results. The postures and movements of the upper arms, trunk, head and wrist are recorded and analysed. The system can be used to assess ergonomic risks, compare different workstations, equipment, and aids, or train work techniques. Previous research has studied the needs and insights of workers and occupational safety and health professionals regarding the use of wearables in the workplace, showing a growing interest of the users and certain concerns regarding data privacy and purpose of use (Jacobs et al., 2019; Schall et al., 2018). However, there is a lack of knowledge of how to present the risk assessment data based on technical measurements for industrial users and stakeholders to support their communication and interpretation of risks. One challenge is that there are few risk assessment methods designed for sensor-based data. Another challenge is that various industrial users, ranging from occupational health and safety (OHS) professionals to production engineers, designers, and managers, may have different perspectives on what information and data are needed on such risk assessment reports. Therefore, in DiPMaS, an iterative design process involving potential industrial user groups is used to ensure that the designed report meets the needs of industrial stakeholders, can support them in interpreting and communicating ergonomic risks, and can be effectively implemented in real working environments.

The objective of this conference contribution is to present the design and evaluation results of three alternative digital ergonomics risk assessment reports based on sensor-measured data of the upper arms, trunk, head and dominant wrist.

Method

The design of the risk assessment report used an iterative process involving industrial stakeholders in several phases (Razzouk & Shute, 2012). In the first phase, a base design was created after a brainstorming session within the research group, which built on previous experience designing ergonomics reports and expertise in this field. In the second phase, the design was presented in group meetings with representatives from four industrial organisations; improvements were made based on the discussions. At this stage, the report includes three sections:

- Part 1: Basic information about the performed measurement and assessed parameters
- Part 2: Exposure and risk assessment results
- Part 3: Summary and, if needed, suggestions for improvements

A complete set of results, including postures and movements of the upper arms, trunk, head and dominant wrist, is provided in the appendix tables.

Three different graphic designs were proposed for Part 2 to visualise the risk assessment results. Traffic light systems were preferred by all industrial representatives and used in all designs. Corresponding graphic illustrations of specific body parts were presented by each assessment item. Design A used modified equal-zone bullet graphs where the background colour zones were of equal size. Design B used a modified bullet graph with a proportional representation of the thresholds. Design C used pie charts showing the time percentage of the specific body parts in each risk zone.

In the third phase, three final designs of the report were evaluated via a digital user survey targeting various industrial stakeholders. The survey included questions from the Interface Usability Instrument (INUIT) on seven factors (Speicher et al., 2015), i.e. *informativeness, understandability, confusion, distraction, readability, information density* and *reachability*, using five-point Likert scales. Additionally, an open question about suggestions was shown after showing each report design. In the final ongoing phase, the ergonomics risk assessment reports will be improved based on the user survey results, and a final evaluation of the report design will be performed.

Results

Fifteen industrial users (five women and ten men) answered the survey in Nov 2024, which were included in this conference contribution; the job roles covered production engineers, designers, managers, and OHS specialists. The total INUIT scores of the report design A, B, and C were 71, 64, and 61 (out of 100 points), respectively. When looking at the seven factors from the INUIT separately, design A showed the best ratings regarding *confusion* and *distraction* of the three designs. Design C had the lowest ratings regarding *information density* compared to the two other designs. Still, all three designs had a median rating of 4 (on a five-point Likert scale) regarding the factors of *informativeness, understandability, readability* and *reachability,* showing that the reports were considered supportive in communicating the ergonomic risks. There were some varying views between the user groups. The open questions provided certain explanation to users' preferences and suggestions for further improvement of the reports.

Discussion and Conclusions

This study used an iterative design process involving industrial stakeholders to design and evaluate ergonomics risk assessment reports based on sensor-measured data. The designed reports can be used to support end users in interpreting and communicating ergonomic risks, comparing and improving the design of workstations or aids, or evaluating work techniques. The methodology ensures active collaboration between researchers and industrial stakeholders to understand user needs, which further guides the development and design of the digital risk assessment report and facilitates the implementation of such digital reports within the industry. The survey showed that the three designs were considered supportive in communicating ergonomics risk assessment results, and design A had the highest rating on all seven factors. In the next stage of the project, the designs of the report will be further improved based on the suggestions from the user survey, implemented in the online digital platform, and tested by users in real work scenarios.

References

Hansson, G.-Å., Balogh, I., Unge Byström, J., Ohlsson, K., Nordander, C., Asterland, P., Sjölander, S., Rylander, L., Winkel, J., Skerfving, S., 2001. Questionnaire versus direct technical measurements in assessing postures and movements of the head, upper back, arms and hands. Scand. J. Work. Environ. Health 27, 30–40. https://doi.org/10.5271/sjweh.584

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. Appl. Ergon. 78, 148–156. https://doi.org/10.1016/j.apergo.2019.03.003

Lim, S., D'Souza, C., 2020. A narrative review on contemporary and emerging uses of inertial sensing in occupational ergonomics. Int. J. Ind. Ergon. 76. https://doi.org/10.1016/j.ergon.2020.102937

Lind, C.M., Abtahi, F., Forsman, M., 2023. Wearable Motion Capture Devices for the Prevention of Work-Related Musculoskeletal Disorders in Ergonomics—An Overview of Current Applications, Challenges, and Future Opportunities. Sensors 23. https://doi.org/10.3390/s23094259

Razzouk, R., Shute, V., 2012. What Is Design Thinking and Why Is It Important? Rev. Educ. Res. 82, 330–348. https://doi.org/10.3102/0034654312457429

Schall, M.C., Sesek, R.F., Cavuoto, L.A., 2018. Barriers to the Adoption of Wearable Sensors in the Workplace: A Survey of Occupational Safety and Health Professionals. Hum. Factors 60, 351–362. https://doi.org/10.1177/0018720817753907

Speicher, M., Both, A., Gaedke, M., 2015. Inuit: The Interface Usability Instrument, in: Marcus, A. (Ed.), Design, User Experience, and Usability: Design Discourse, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 256–268. https://doi.org/10.1007/978-3-319-20886-2_25

Takala, E.-P.P., Pehkonen, I., Forsman, M., Hansson, G.-A.Å., Mathiassen, S.E., Neumann, W.P., Sjøgaard, G., Veiersted, K.B., Westgaard, R.H., Winkel, J., 2010. Systematic evaluation of observational methods assessing biomechanical exposures at work. Scand. J. Work. Environ. Health 36, 3–24. https://doi.org/10.5271/sjweh.2876

Gender in human factors education: A pilot study

Rich C. McIlroy

Transportation Research Group, University of Southampton

SUMMARY

Women are chronically underrepresented in engineering. Some have argued that greater emphasis of the social relevance of engineering could help address the challenge. Human factors and ergonomics (HFE), in its focus on the human in the system, does just this. Could HFE therefore help us solve the engineering gender challenge? How is HFE currently performing in terms of gender equity? This study used surveys and interviews with students and educators to explore these issues. Participants highlighted the need for institutional guidelines, sincerity in implementation, and a broader, integrated approach. HFE, by virtue of its human-centred focus, is well-positioned to lead efforts toward gender equity in engineering, but concerted effort is required.

KEYWORDS

Human factors and ergonomics education; gender equity; engineering education

Introduction

Engineering fields suffer from stark gender imbalances, with women comprising only 31% of science, technology, engineering, and maths (STEM) university students in the UK (HESA, 2023). Although the 'men are technical and women are social' dichotomy is a socially constructed and structurally perpetuated fallacy, it nevertheless holds true that incorporating more human-centric perspectives into engineering education could attract a more women into the domain (Faulkner, 2007; Nilsson, 2015). Human factors and ergonomics (HFE) as a sub-discipline that is concerned with designing systems for human use, inherently integrates social and technical elements. It could therefore play a key role in fostering gender equity in engineering; however, little research has addressed HFE's educational potential in this context.

Gender mainstreaming in education, which embeds gender considerations across curricula, has been identified as a strategy to address disparities (Peña & de les Valls, 2023); however, there is resistance to this approach (Verge et al., 2018). Incorporating more HFE content into engineering education could circumvent resistance and contribute to greater gender equity. It would also contribute to better engineers (Boiron et al., 2022). This research seeks to explore these issues, asking: to what extent are requirements for gender mainstreaming in engineering met in HFE education specifically, what is the status quo, and to what extent does HFE represent a suitable vehicle for greater incorporation of the gender perspective into engineering education?

Method

The study combined surveys and semi-structured interviews. The questionnaire included open-ended questions asking participants' opinions on the extent to which gender is incorporated into engineering and HFE education. It also included a question on participants' views of the humanisation of engineering as a route to gender equity, and the role that HFE education could play therein. Respondents were also asked about the extent to which 24 specific aspects of gender mainstreaming were already incorporated into engineering and human factors education, with responses invited on a six-point Likert scale from 'Not at all' to 'Completely' (adapted from Peña et al., 2021). The follow-up interviews provided deeper context, further insight into experiences, and details on institutional support for gender-focused teaching. The study was approved by the University of Southampton ethics committee (ID 89859).

Results

Questionnaire: Participants

A total of 21 individuals provided sufficiently complete responses to the questionnaire, 17 of whom were female, three of whom were male, one of whom self-described as gender fluid. One male did not provide his age, the others were between 23 and 81 years of age. The females' ages ranged from 22 to 63 (mean = 35.5, SD = 11.4), and the gender fluid individual was 21. Ten respondents gained their experience in the UK, four in Australia, two in the US, and one in Canada. Four did not indicate where they had gained their HFE experience. Results regarding the respondents' level of experience and the recency of that experience are presented in Figures 1 and 2.

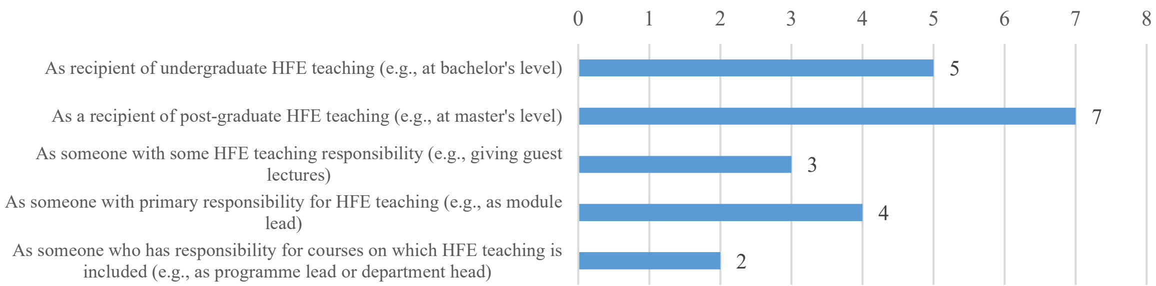


Figure 1. Respondents' level or types of HF experience

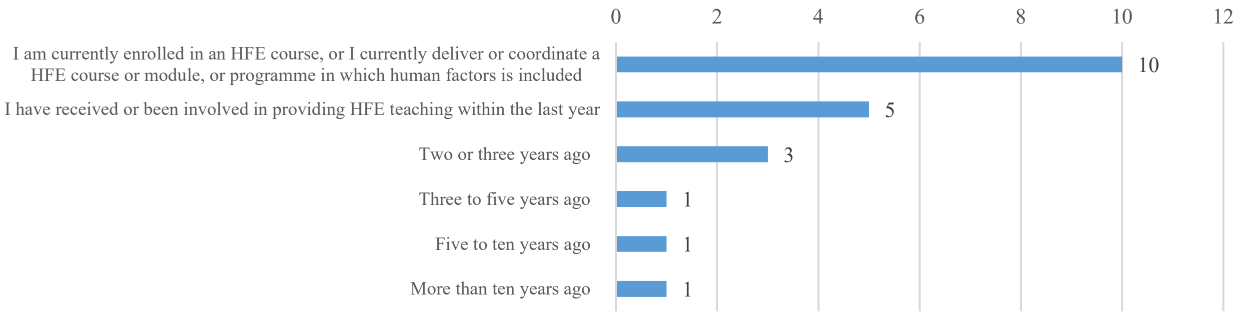


Figure 2. Recency of respondents' experience

Questionnaire: quantitative findings

Figure 3, below, shows the distribution of responses to the two questions asking respondents about the extent to which they thought gender was incorporated into engineering education and into HFE education, on a scale of one (not at all incorporated) to 10 (completely incorporated). Taken as a group, the respondents were more positive about the extent to which gender was included in HFE education compared to engineering more broadly.

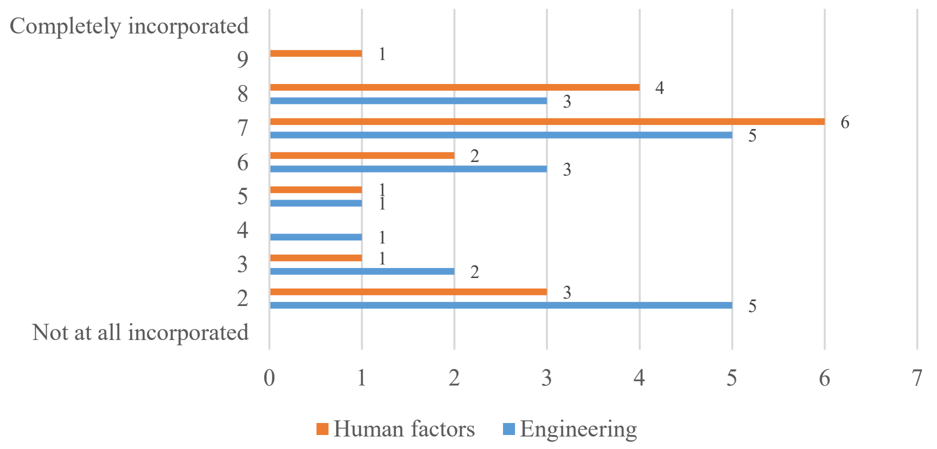


Figure 3. Distribution of responses to the questions “To what extent do you think the sex and gender perspective is incorporated into current engineering / human factors and ergonomics teaching at your institution?”

Table 1, below, displays all items of the section of the questionnaire taken from (Peña et al., 2021) on the methods by which gender could or should be incorporated into education. Responses to the question *“In your experience of human factors and ergonomics education, to what extent are the following aspects incorporated into or included in teaching at your institution?”* were invited on a six-point scale from not at all (one) to completely (six). Table 1 provides the means, standard deviations, and ranges of responses to each item.

Finally, Figure 4 displays the distribution of responses to the question *“If you have responded to this questionnaire as an educator, do you feel ready, sufficiently equipped, and/or supported internally or externally to start incorporating the gender perspective into your teaching and/or module/course design activities?”*. Note that this question was only presented to those that had indicated being an educator, not only a student, of HFE. As can be seen, none of the 11 individuals answering this question said that they already fully incorporated the gender perspective into their teaching. One said that they did not see the need to.

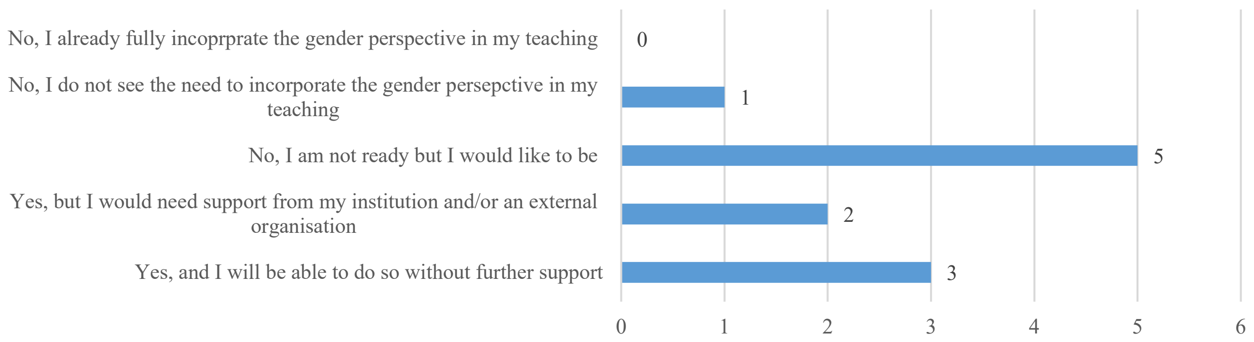


Figure 4. Distribution of responses to the item asking educators the extent to which they felt ready, sufficiently equipped, and/or supported internally or externally to start incorporating the gender perspective into their teaching activities (n = 11).

Table 1. Mean responses to the items taken from (Peña et al., 2021) on the ways in which gender can be incorporated into education, recorded on a scale from one (not incorporated at all) to six (completely incorporated).

Item	Mean	SD	Range
Explaining the usefulness of the gender perspective	3.2	1.4	1-6
Explaining the social relevance of the gender perspective	2.9	1.5	1-5
Including references of female authors and/or professional women	3.8	1.7	1-6
Including the full first names of authors referenced	3.1	1.9	1-6
Ensuring images used in teaching do not perpetuate gender stereotypes	3.6	1.7	1-6
Ensuring the use of non-sexist and non-androcentric language in teaching materials	4.0	1.6	1-6
Using examples in class that challenge or counter gender stereotypes	3.2	1.7	1-6
Ensuring examples and exercises cover various topics	4.5	1.3	1-6
Allowing students to participate in defining some of the content	3.8	1.6	1-6
Ensuring learning is active, combining participatory activities, case studies, projects, etc.	5.1	1.2	3-6
Promote the study of an issue that has social and/or gender relevance	3.5	1.7	1-6
Highlighting where a woman plays a relevant role in a case study	3.0	1.7	1-6
Including at least one lecture or video delivered by a woman	4.4	1.5	1-6
Explicitly highlighting social and/or gender relevance in teaching activities	2.9	1.5	1-6
Giving guidelines on inclusive language for presentations and/or reports	2.8	1.4	1-6
Holding in-class discussions on aspects related to the gender perspective	2.5	1.5	1-6
Analysing potential gender imbalance in in-class participation	2.7	1.4	1-5
Protecting female participation from (male) interruption	3.1	1.7	1-6
Ensuring students' responses are treated with respect and tolerance	5.2	.83	3-6
Ensuring students use non-sexist, inclusive language in class	4.5	1.3	1-6
Analysing the distribution of roles in group work and discuss in class	3.3	1.6	1-6
Promoting rotating roles in group work	3.5	1.8	1-6
Empowering females in work groups by ensuring they are addressed, listened to, etc.	3.2	1.5	1-6
Ensuring the educator is accessible outside of class hours to address student queries / concerns	4.9	1.2	2-6

Questionnaire: Qualitative results

Free text responses were provided to three questions. The first concerned the gender perspective in engineering education, asking: "*If you think it is not sufficiently or completely incorporated, or that more could be done, what else do you think could be included or done? Why, or in what way? If you think it is sufficiently incorporated, could you give examples of how it is currently done (e.g., best practices), or why you think further inclusion unnecessary (if that is your view)?"* The second question was worded in the same way but asked about HFE education rather than engineering education. The third question asked about the humanisation of engineering, stating: *"Greater humanisation of engineering education has been argued to represent a potential route to greater gender equity in STEM education (science, technology, engineering, maths) (e.g., see: https://doi.org/10.1007/s10668-023-03667-2). Considering this, how do you see the role of human factors and ergonomics in contributing to gender equity in broader engineering education and beyond (e.g., in the workplace or research domain)?"*.

Gender in engineering education

Respondents provided a total of 622 words in response to this question, across 14 responses (with an average of 44.4 words per response, and a range of 12 to 99 words). Three individuals referred to a lack of attention paid to gender in engineering education or the acknowledgement that it is or should be an issue of importance (e.g., *"It's not mentioned and rarely acknowledged as a significant* *factor"* female, 43). Where respondents discussed its inclusion, this was typically in relation to HFE teaching (or a related topic), with some discussion of physical ergonomics or anthropometrics as the main topic via which sex and/or gender is covered (e.g., *"If it does happen to be mentioned, it is usually associated with anthropometrics"* female, 32); however, caveats were often provided (e.g., *"In terms of gender equity, the teaching has some inclusion of human centred design but I wouldn't say there is explicit teaching on gender perspectives",* female, 42).

Interestingly, and in advance of a question a related question that appeared later in the questionnaire, one respondent indicated that *"Human factors and ergonomics is typically not sufficiently incorporated into engineering teaching within my discipline"* (female, 32), and another discussed a lack of support from their faculty: *"I haven't overtly included a sex or gender perspective in my teaching and haven't had guidance from my institution to do so"* (female, 39).

Several respondents discussed aspects that went beyond the content of teaching, including the presence of female staff and the impact of having a woman in an influential faculty position (e.g., *"We have the first female dean and she is very worried about promoting women in science"*, female, 41) and the consideration for learning and interaction styles in class activities (*"It is mostly ignored especially in modules that require group work who just ignore the way that gender can influence that experience"*, female, 22).

Gender in HFE education

Respondents provided a total of 415 words in response to this question, across 9 responses (with an average of 51.9 words per response, and a range of 12 to 104 words). Fewer responses were provided to this question, perhaps in part due to its similarity to the previous question and the finding that several people discussed HFE teaching when asked about engineering (and may not have felt the need to repeat themselves). Of the nine respondents that provided responses to this item, five expressed what could be considered as negative sentiments, e.g., *"It's not mentioned at all that perspectives could be different or varied"* (female, 43).

Two respondents were more positive, discussing how the issue is currently incorporated into the teaching of which they had experience: *"In the HF/E course there is acknowledgment of HCD* [human centred design] *to accommodate for diversity"* (female, 42). Another respondent discussed more generally how gender is inherent to the HFE approach: *"In my opinion, human factors provides knowledge and tools to work with human variability. This includes sex, age, disabilities, etc. This perspective is always present."* (female, 41).

How do you see the role of HFE in contributing to gender equity?

Respondents provided a total of 626 words in response to this question, across 13 responses (with an average of 48.1 words per response, and a range of 8 to 110 words). Many of the responses to this question centred on HF being a potential leader in the strive for greater gender equity in engineering: e.g., *"I think HFE has a vital role to play … given its underpinning humanistic and participatory roots"* female, 41. Some expressed sentiments more related to the content of teaching (e.g., *"It certainly highlights how different genders interact with engineering in different ways"* female, 24), whereas others referred more to other contributions that could be made (e.g., *"By studying HFE and talking about the issues that occur groups may feel less ignored or marginalised which in turn increases engagement in the engineering disciplines as a whole"* female, 22).

One respondent made a point about the relevance of HFE, the greater proportion of women in the field compared to other engineering disciplines, and the potential impact on gender equity in broader engineering: *"I know that there is a higher proportion of women in HFE compared to other engineering disciplines and STEM fields more broadly. I think that HFE will continue to grow in relevance…This may make women in STEM more visible in companies"* (female, 24). That said, a warning from another respondent highlighted the importance of the ways in which HFE teaching is incorporated: *"Educators must understand WHY there is a need for this, so students can be convinced that this is not just a fashionable thing to do or a box-ticking exercise"* (female, 39). Another respondent commented on the difficulty that could be faced in such an exercise: *"It's hard enough even getting the HF 'voice' adequately respected in engineering (and this is from someone who is a HF lecturer and is also a woman engineer), let alone gender perspectives"* (female, 42).

In contrast to these positive comments, one respondent indicated disagreement with the need to emphasise the gender perspective, stating: *"I personally disagree with highlighting gender issues over others, such as aging or disability, which are also very relevant"* (female, 41). This should be considered alongside the warning that implementation requires careful thought.

Finally, although the current research specifically focussed on teaching rather than research (which has begun to be addressed elsewhere: Parnell & Plant, 2024; Parnell et al., 2022; Read et al., 2022), one respondent mentioned its importance for education: *"although this survey is focussed on education not research it is important to support gender equity in research so that there are good opportunities to cite women scholars in teaching materials on key HFE topics"* (female, 41).

Interviews

Five survey respondents indicated being interested in participating in a follow-up interview by providing their email address at the end of the questionnaire. Three of those responded to emails and were interviewed, all of whom were women. Two were from the same institution, one of whom was a 33-year-old PhD student with seven years of experience in maritime engineering, the other a 37-year-old associate professor with 15 years of experience in human factors. The third was a 39-year-old assistant professor with 20 years of experience in human factors, user interface design, and human-computer interaction. The latter two had received and delivered HFE teaching at the university level, the PhD student had taken a module in HFE as part of her undergraduate education and was incorporating HFE into her PhD. All came from institutions in which HFE teaching is embedded in an engineering faculty. Each interview lasted approximately 30 minutes.

Each interviewee was asked whether they considered gender to be relevant in engineering education and to what extent it was already covered. The assistant professor indicated that *"it is not really talked about"*, though went on to describe an example where a PhD student of hers had been *"mansplained by one of the lecturers who thought he was doing the right thing"* when being shown examples of female engineers that students should be aspiring to. Considering the inclusion of the items 'Including references of female authors and/or professional women' and 'Highlighting where a woman plays a relevant role in a case study' in the guidelines provided by (Peña et al., 2021) (in Table 1), this highlights a danger, with the method of implementation (alongside sincerity) crucial if resistance is to be avoided.

The associate professor noted a top-down, imposed need to include equity considerations in teaching, commenting that it *"feels like there's a bit of reverse engineering going on"*. She was, however, largely positive, intimating that *"in the last couple of years, it's become much more considered"*. The PhD student was also positive (*"as time has passed, it's becoming more prevalent in engineering education"*); however, she noted that the education system is *"still not considering equity in and of itself and it's not considering gender specifically"*. When asked specifically about HFE education, all three stated that gender is already incorporated. The associate professor referred to demographics in human-based research, the assistant professor discussed representative data, and the PhD student discussed how gender *"is fundamental to human factors"*.

The two educators were asked if they had received guidance on incorporating gender into teaching, or why this might be useful, to which both answered no. Where they had incorporated the topic, they discussed personal rather than institutional motivations. For example, the associate professor discussed her experiences of becoming a mother (*"I didn't really ever think it was a big issue until I had children and then I really could see…no one's ever told me or taught me about it"*).

To explore how the interviewees thought gender might be better integrated, the questionnaire prompts displayed in Table 1 (above) were shared on screen and discussed. Both the PhD student and the associate professor discussed gender-neutral language as a key issue, e.g., *"I do a lot in, like unmanned aerial vehicles, like UAVs, and I always say crewed, un-crewed, to you know, very senior professors. It's just a natural 'unmanned' and I always say un-crewed"* (associate professor). The most salient message, however, was that there is currently no top-down support for educators.

The interviewees were asked which they thought would be more suitable, a transversal approach whereby gender is incorporated into all aspects of engineering education, or a focussed approach whereby it is provided as part of a specific subject (as discussed by González-González et al., 2020). The associate professor articulated the view that a transversal approach would be more suitable: *"I think it's that classic like little and often, if you just have it in one module, people might switch off more easily or think oh, it's not really relevant to me"*. The PhD student went further to state *"I would lean towards a transversal approach for integrating human factors education into engineering as a whole as well"*, suggesting that incorporating human factors across engineering modules represents a means for incorporating gender, and other social aspects, into engineering.

Discussion

This pilot study has sought to present some initial discussions around gender in human factors and ergonomics (HFE) education and of the role of HFE in the socialisation and gender equitisation of engineering education at the tertiary level. In terms of the specific tools and methods by which engineering gender inequity might be addressed (Peña et al., 2021) (see Table 1), results paint a mixed picture. Active and participatory learning activities, ensuring the educator is available outside of class hours, and respecting students' in-class responses are all well implemented. These are markers for high quality, engaging teaching in a general sense. In contrast, gender specific factors such as analysing in-class gender imbalances, holding discussions related to gender, and highlighting the social relevance of gender, are not well implemented. This implies a well-established approach to engaging, respectful teaching, but a lack of directed attention to gender. One might suggest Advance HE's Athena Swan framework (Advance HE, n.d.-a) to go some way to address this; however, the practical benefit and sincerity of Athen Swan implementation was questioned by the associate professor interviewee: *"I just found it all a bit tick, boxy. At the moment, there's so much, you know, Athena Swan this and that, and it's … hard to stomach sometimes"*. This is still clearly a challenge, with performative policies a major barrier to true equality in higher education (Smidt et al., 2021).

The results suggest that the gender gap in STEM is not being sufficiently addressed. Although participants were more positive about HFE compared to wider engineering, most did not consider the issue of gender to be sufficiently addressed in either. Where it is considered, many spoke of a superficial treatment majoring on anthropometrics. Furthermore, recognition of shortcomings in data and physical systems is key (e.g., Criado Perez, 2019), but insufficient.

Despite the existence of guidelines that mandate the inclusion of equity issues in engineering (Engineering Council, 2020), none of the participants of this research was wholly satisfied with how it is currently covered. Although HFE sees a more equal gender balance than most engineering sub-disciplines (Lum et al., 2022), and respondents were somewhat more positive about HFE compared to engineering more broadly, more could certainly be done. For example, results suggest there to be little institutional support for direct treatment of gender by university educators, despite a relatively long history of effort in this domain (e.g., Cronin & Roger, 1999). Incorporating the factors outlined by Peña et al. (2021) into training programmes targeted at early career lecturers, for example through Advance HE's widely adopted fellowship programme (Advance HE, n.d.-b), could represent a suitable initial avenue for this. That said, this may not go far enough: *"I think PGCAP* [post-graduate certificate in academic practice] *will capture a certain demographic of … newer younger academics… If it's going to be done properly, there needs to be some kind of best practise guidance"* (associate professor).

The participants, particularly the interviewees, suggested HFE to represent an ideal vehicle for the greater incorporation of the social relevance aspect into engineering. The perceived societal relevance of engineering is one of the most crucial factors for increasing women's participation in the field (Baker et al., 2007), and acknowledging the societal relevance of engineering concepts has been highlighted as necessary for enhancing and sustaining young women's motivation to pursue engineering (Godwin et al., 2016). This needs to be capitalised on by the HFE community. There was broad agreement that a transversal approach, whereby issues are integrated across subjects and modules, rather than a focussed approach, in which gender is focussed upon in a single subject or module, is required. This is already happening in engineering education in relation to the UN's sustainable development goals (SDG) (Pérez-Foguet & Lazzarini, 2019). Notably, one of those goals is gender equality (United Nations, n.d.), yet there remains a long-standing resistance to gender mainstreaming in education (Acker, 1988; Tildesley, 2023; Verge et al., 2018). Therefore, taking a transversal approach to integrating HFE material across engineering sub-disciplines, and including gender as a core topic within that integration, could present a way to embed this SDG whilst overcoming some of the observed resistance; however, such change would need to be driven by the institution rather than by individual educators (Peterson & Jordansson, 2022). Moreover, although this would contribute to greater focus on the social benefit of engineering and provide a stronger end user perspective in teaching content, it would not necessarily address the teaching style factors outlined by Peña and colleagues (Peña & de les Valls, 2023; Peña et al., 2021). These would still apply to any higher education teaching activities, in HFE or other engineering sub-disciplines, or indeed any other subject within and beyond the STEM field.

Limitations

Given the small sample sizes, results must be treated with caution. This was a pilot study to support initial explorations in the field of gender, human factors, and engineering, hence is not positioned as a full exploration of themes. Much more attention in this field is required, with engagement from participants from a broader range of institutions and greater international representation required to provide more detailed and robust answers to the research question posed above. It is also important to highlight the gender imbalance in the sample, with most questionnaire respondents and all interviewees being women. The path to societal gender equity cannot only involve women (Kimmel, 2005). This is true broadly as well as specifically in the context of engineering (Wilson et al., 2021). Greater male participation in this kind of research is therefore crucial.

The qualitative nature of this research requires recognition of the biases of the analyst. I am a white, cis-gender, male researcher attempting to approach the issue as an ally, from a belief that a more gender equitable society will benefit all genders. Had this research been undertaken by a woman, or someone that does not conform to the binary male/female distinction, different insights (and indeed different data, particularly interview data) would likely have resulted. Although this does not diminish the utility of this study, it does merit explicit acknowledgement.

Conclusions

This discussion piece and pilot study has explored the potential for greater integration of HFE into engineering education at the university level to help address the long-standing and persistent gender imbalance seen in STEM. Emphasising the societal relevance of engineering has been acknowledged as a way to encourage more women into the field. HFE, as an engineering sub-discipline, addresses this point, hence offers a natural vehicle for gender-focussed teaching strategies and content. That said, care must be taken with implementation to avoid resistance, and educators must have guidance to draw on.

References

Acker, S. (1988). Teachers, gender and resistance. *British journal of sociology of education*, *9*(3), 307-322.

Advance HE. (n.d.-a). *Athena Swan Charter*. Retrieved 03/07/2024 from https://www.advance-he.ac.uk/equality-charters/athena-swan-charter

Advance HE. (n.d.-b). *Fellowship*. Retrieved 04/07/2024 from https://www.advance-he.ac.uk/fellowship/fellowship

Baker, D., Krause, S., Yaşar, Ş., Roberts, C., & Robinson-Kurpius, S. (2007). An intervention to address gender issues in a course on design, engineering, and technology for science educators. *Journal of engineering education*, *96*(3), 213-226.

Boiron, O., Deumié, C., Raviol, L., & Benech-Kopelianskis, M. (2022). Incorporating the Gender Perspective in Engineering Curricula: The Case of École Centrale Marseille. In *Overcoming the Challenge of Structural Change in Research Organisations–A Reflexive Approach to Gender Equality* (pp. 143-157). Emerald Publishing Limited.

Criado Perez, C. (2019). *Invisible women: Data bias in a world designed for men*. Abrams.

Cronin, C., & Roger, A. (1999). Theorizing progress: Women in science, engineering, and technology in higher education. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *36*(6), 637-661.

Engineering Council. (2020). *The Accreditation of Higher Education Programmes (AHEP) Fourth Edition*.

Faulkner, W. (2007). `Nuts and Bolts and People':Gender-Troubled Engineering Identities. *Social Studies of Science*, *37*(3), 331-356. https://doi.org/10.1177/0306312706072175

Godwin, A., Potvin, G., Hazari, Z., & Lock, R. (2016). Identity, critical agency, and engineering: An affective model for predicting engineering as a career choice. *Journal of engineering education*, *105*(2), 312-340.

González-González, C. S., García-Holgado, A., & Garcia-Peñalvo, F. J. (2020, 27-30 April 2020). Strategies to introduce gender perspective in Engineering studies: a proposal based on self-diagnosis. 2020 IEEE Global Engineering Education Conference (EDUCON),

HESA. (2023). *What do HE students study?* Retrieved 13/05/2024 from https://www.hesa.ac.uk/data-and-analysis/students/what-study#complete

Kimmel, M. S. (2005). Why men should support gender equity. *Women's Studies*, *103*, 102-114.

Lum, H. C., Grier, R., Waldfogle, G., Hancock, G. M., Lerner Papautsky, E., & Hughes, A. M. (2022). Reflections on Gender Bias and Disparity in the HF/E Profession. Proceedings of the Human Factors and Ergonomics Society Annual Meeting,

Nilsson, L. (2015, 27th of April, 2015). How to attract female engineers. *The New York Times*.

Parnell, K. J., & Plant, K. L. (2024). How can human factors close the gender data gap? *Human Factors and Ergonomics in Manufacturing & Service Industries*, *34*(1), 63-75.

Parnell, K. J., Pope, K. A., Hart, S., Sturgess, E., Hayward, R., Leonard, P., & Madeira-Revell, K. (2022). 'It's a man's world': a gender-equitable scoping review of gender, transportation, and work. *Ergonomics*, *65*(11), 1537-1553. https://doi.org/10.1080/00140139.2022.2070662

Peña, M., & de les Valls, E. M. (2023). Inclusion of the gender equality sustainable development goal in engineering teaching and research. *Environment, Development and Sustainability*. https://doi.org/10.1007/s10668-023-03667-2

Peña, M., Olmedo-Torre, N., Mas de les Valls, E., & Lusa, A. (2021). Introducing and Evaluating the Effective Inclusion of Gender Dimension in STEM Higher Education. *Sustainability*, *13*(9), 4994. https://www.mdpi.com/2071-1050/13/9/4994

Pérez-Foguet, A., & Lazzarini, B. (2019). Continuing professional education in engineering faculties: Transversal integration of sustainable human development in basic engineering sciences courses. *Journal of Cleaner Production*, *218*, 772-781. https://doi.org/https://doi.org/10.1016/j.jclepro.2019.02.054

Peterson, H., & Jordansson, B. (2022). Gender mainstreaming in Swedish academia: translating policy into practice. *Journal of Gender Studies*, *31*(1), 87-100.

Read, G., Madeira-Revell, K., Parnell, K., Lockton, D., & Salmon, P. (2022). Using human factors and ergonomics methods to challenge the status quo: Designing for gender equitable research outcomes. *Applied ergonomics*, *99*, 103634.

Smidt, T. B., Pétursdóttir, G. M., & Einarsdóttir, Þ. (2021). When discourse is hijacked: An implicit and performative resistance strategy to gender equality in higher education. *Journal of Women and Gender in Higher Education*, *14*(2), 143-165.

Tildesley, R. (2023). Transforming academic research? Resistances to gender mainstreaming implementation in universities. *European Journal of Women's Studies*, *30*(4), 486-501.

United Nations. (n.d.). *Goal 5: Achieve gender equality and empower all women and girls*. Retrieved 04/07/2024 from https://www.un.org/sustainabledevelopment/gender-equality/

Verge, T., Ferrer-Fons, M., & González, M. J. (2018). Resistance to mainstreaming gender into the higher education curriculum. *European Journal of Women's Studies*, *25*(1), 86-101.

Wilson, N. L., Dance, T., Pei, W., Sanders, R. S., & Ulrich, A. C. (2021). Learning, experiences, and actions towards advancing gender equity in engineering as aspiring men's allyship group. *The Canadian Journal of Chemical Engineering*, *99*(10), 2124-2137.

How well can generative AI design and evaluate user interfaces?

Zhenyuan Sun and Chris Baber

University of Birmingham

SUMMARY

The inexorable rise of generative artificial intelligence (GenAI) is threatening a range of work domains. In this paper we explore whether user interfaces produced by GenAI compare with those produce by humans, and whether gen-AI can evaluate user interfaces to a human-like standard. We create user interface designs for a burger ordering app using prompts to Midjourney on Discord, DALL-E 3 on ChatGPT4o, and Stable Diffusion 3 on Stable Assistant. All three GenAI apps had problems with legible text and following prompts provided. However, through adjustment of prompting, DALL-E 3 and Stable Diffusion 3 produced viable designs that met the brief. We compared the resulting designs with commercial products and the designs created by 8 competent (human) user interface designers through a survey of 32 participants evaluating the designs using the UEQ-S. We found no difference in pragmatic quality between designs, but the designs from gen-AI were rated significantly higher on hedonic quality than those from the commercial products or human designers (with the commercial apps having lowest ratings on all measures). We then prompted ChatGPT4o and Stable Assistant using the UEQ-S to evaluate the user interface designs. We found little correlation between the ratings of the gen-AI apps and human raters. This suggests that GenAI might have a place (with appropriate prompt engineering) in generating user interface designs but that, at present, it struggles to produce reliable, human-like evaluation of these.

KEYWORDS

Generative AI; user interface design; user interface evaluation

Introduction

Generative Artificial Intelligence (GenAI) is capable of creating (or synthesising) novel output. Image generation models, such as DALL-E, Midjourney, Stable Diffusion, Dream Studio, Amazon Nova, are capable of creating images in response to text prompts. The manner in which the images are created can be complicated, but the basic approach is one of synthesising material scraped from the World Wide Web. Such scraping of data has led to significant (and unresolved) concerns over Intellectual Property Rights[1], although the synthesising processes are sufficient stochastic to imply that it is parts of (rather than complete) images that are being used. More broadly, there is an ongoing debate over whether genAI might be more creative than humans. On some measures, such as divergent thinking tests or association tasks, GenAI produces superior performance (Cropley, 2023), although this had some caveats that suggested that the genAI was not producing highly novel solutions, and it remains that case that the best humans are more creative than AI (Koivisto and Grassini, 2023). That AI can even be considered as challenging creative jobs had led to much concern and debate about the type of work that might be at threat (Lee, 2022; Aliyev, 2023).

[1] https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem [accessed 27/01/25]

In this study, we use GenAI to design to user interfaces and perform User Experience evaluation. While there is, at present, no agreed approach to the use of GenAI in user interface design, it might be useful for initial prototype sketching, particularly for people with little or no prior experience in user interface design. In such a use-case, the user might describe the user interface they wish to develop and ask GenAI to provide some examples. This, of course, speaks directly to the threats to creative work because the initial concept design of user interfaces is not simply a matter of making drawings to hand over to a software developer. Rather, one would expect effort to be put into the capturing of user requirements. In our example, we assume that requirements have already been collected (although, of course, it is easy to define a prompt that might define a set of requirements from fictitious users, Ronanki et al., 2024.).

Having defined user interface designs, we consider how well GenAI could perform user evaluation on these. Large Language Models (LLMs) are proving popular in synthesizing human data (Argyle2023; Aher2023;Park2023) and have produced impressive results across a range of domains. However, Gui and Toubia (2023) urge caution on becoming overly reliant on the approach. They explored the use of LLMs to generate consumer decisions and showed problems in the generated data. They suggest that these problems might be reduced somewhat through prompt-engineering (e.g., adding detail about customers and their environment might help manage covariates, and providing information to the LLM on the purpose of the analysis can lead to more plausible outcomes). Even with these measures, Gui and Toubia (2023) suggest that there are likely to be outputs from the LLM that ignore the cause-effect relations that are implied in the prompt. This need not simply be due to 'hallucination' but could arise from plausible (but erroneous) inferences that the LLM draws from the content of the prompt.

In the context of user experience evaluation, Pourasad and Maalej (2024) demonstrate that, given a prompt from which the GenAI is asked to take on the role of user interface evaluator and report usability problems, it is possible for the genAI to identify *some* of the problems that were identified by either usability testing or expert evaluation. This is summarised in table 1. This suggests that expert review identified around 50% of the possible issues, but the genAI identified a similar proportion to usability testing. Given that usability testing is often considered an expense, one can see the temptation to replace this with an automated testing regime.

Table 1: Summary of Data from Pourasad and Maalej (2024)

Approach	Usability issues (total 110)	Unique issues identified by approach
Usability testing (with 10 users)	25	8
Expert review (with 2 experts)	54	31
genAI (with UX-LLM)	30	8

In the study reported in this paper, we first use different GenAI models to design user interfaces for a mobile app, and then use two genAI tools to evaluate these design using the User Experience Questionnaire.

Method

As noted in the introduction, we have assumed that user requirements have been captured. We also wanted to use an example for our design exercise that had already been reported, so that we could consider what would be acceptable as a 'good' solution to the exercise. We used the example of

‘Tasty Burger’ which Yalansaka et al[2]. (2018) have presented as an exercise in user interface design from requirements gathering to specification to design sketches.

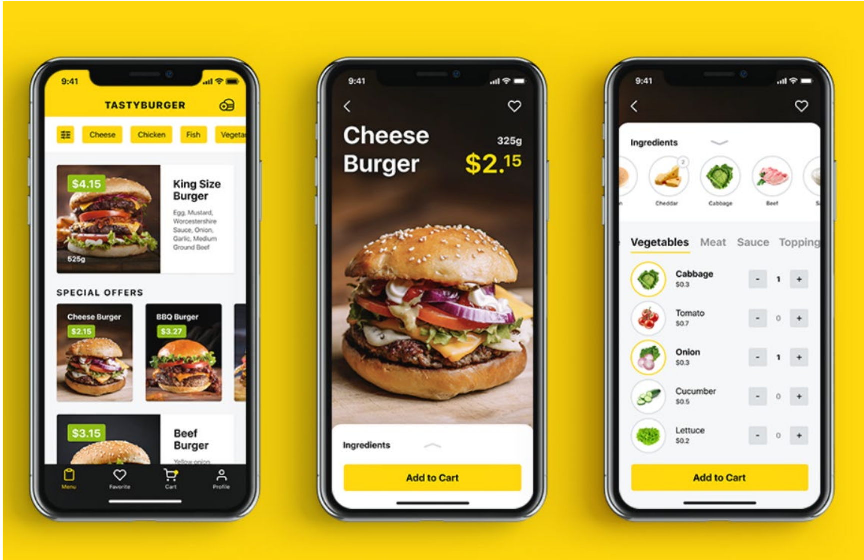


Figure 1: ‘Tasty Burger’ design exercise user interface designs

User Design Exercise

From the ‘tasty burger’ design exercise, we defined a design brief that summarised user requirements and design considerations. Five requirements were defined:

1. Present users with the basic data about the items such as photos, ingredients, and weight.
2. Allow customers to find a needed item quickly.
3. Allow hungry new users to understand how the app works quickly.
4. Allow users to customize any option of the burgers by adding or removing the ingredients from a burger.
5. Be easy to use for customers who buy an item regularly.

The designers were asked to create a high-fidelity prototype of the app’s interface based on these requirements. Since three pages of the app were presented in Figure 1, the designers were also required to limit their design to three pages. The designers could assign the contents of the three pages as they wish, but they needed to make sure that their pages illustrate functions that meet the requirements. In addition, human designers were free to use any graphics editors they liked, but were forbidden to use any image-generative AI functions of these kinds of software or any other AI programs.

We gave the brief to 5 students on the MSc Human-Computer Interaction degree programme at the University of Birmingham and asked them to produce concept designs. The concept designs were created using Figma (3), Sketch (1), or JS Design (1). In addition to producing user interfaces, we also asked them to explain the user journey for their designs.

GenAI Design Exercise

The brief was also presented to three gen-AI apps: Midjourney on Discord, DALL-E 3 on ChatGPT4o, and Stable Diffusion 3 on Stable Assistant. In order not to unduly influence the GenAI

[2] Yalanska, M., Morozov, A., Asanov, E. and Taran, V., 2018, Case study: Tasty Burger. UI design for a food ordering app, https://blog.tubikstudio.com/case-study-tasty-burger-ui-design-for-food-ordering-app/

design, we defined a prompt that was as close to the brief to the participants. This is shown in the text box below:

> Text-to-Image:
>
> You are a user interface designer and are creating the user interface for a mobile app.
>
> Generate a novel 3-page user interface design of a burger-ordering app that allows hungry customers who do not want to devote much time and effort to learn how an app works to easily order a traditional burger from the menu or after customizing any options of the burger.
>
> The user interface design should:
>
> 1. Present users with the basic data about the items such as photos, ingredients, and weight.
> 2. Allow customers to find a needed item quickly.
> 3. Allow hungry new users to understand how the app works quickly.
> 4. Allow users to customize any option of the burgers by adding or removing the ingredients from a burger.
> 5. Be easy to use for customers who buy an item regularly.

Initial Observations on genAI activity

Each GenAI had difficulty in generating readable text or following the design requirements. While modification of the prompts improved the viability of designs, there were still issues that required addressing.

In DALL-E 3 on ChatGPT 4o, the text rendered in the images was unreadable. This could be due to the genAI not treating the text as words but merely as image features. When text was readable, there were problems including misspelt words, nonsense phrases, and words made up of illegible characters. In addition, even when words were correctly spelt and rendered in the image, there were mistakes in their placements, such as duplicate presentation in a single row. Also, some graphic elements had issues of plausibility, such as oddly shaped burgers, unrecognizable ingredient pictures, and some graphic elements repeatedly appeared on a single page, such as on the menu.

Midjourney, run through Discord, had problems with generating the required number of pages (often generating 4 or more pages, despite the prompt). Midjourney also had problems with text and images. In addition, a substantial proportion of images created by Midjourney were presented from a non-vertical viewing angle and some UI pages overlapped, which made it difficult to recognize some details of the design. A further issue with Midjourney was that it does not have the function to describe its design of user flow, as DALL-E 3 can do through ChatGPT, which made it difficult to use additional prompts to refine the generated image. For this reason, Midjourney's design was not adopted for the next part of the study.

Stable Diffusion 3 on Stable Assistant had problems with number of pages, incomplete images, and non-vertical viewing angles. Additionally, Stable Diffusion 3 sometimes presented the UI of mobile apps embodied in a phone with exaggerated length.

We responded to these problems through running the GenAI process multiple times (i.e., for Stable Diffusion 21 runs were required to produce five usable designs) and then selecting user interface designs that met the study requirements.

Evaluation by Human Raters

To produce a comparable sample, we selected 5 sets of images for each GenAI (because we had 5 HCIStudents create designs). We also used 5 designs from commercial organisations. This gives a total of 20 sets of 3 screens to evaluate: 5 designs each from HCIStudents, Dall-e 3 and Stable Diffusion 3, and designs from commercial organisations (i.e., Burger King, McDonald's, KFC, Five Guys). The commercial user interfaces were redrawn by one of the authors (using Photoshop) to ensure consistency of image size. As far as possible, the images were kept as close as possible to those generated by GenAI or designed by participants. For the commercial organisations, identifying logos and product names were substituted with the names used in the 'tasty burger' app.

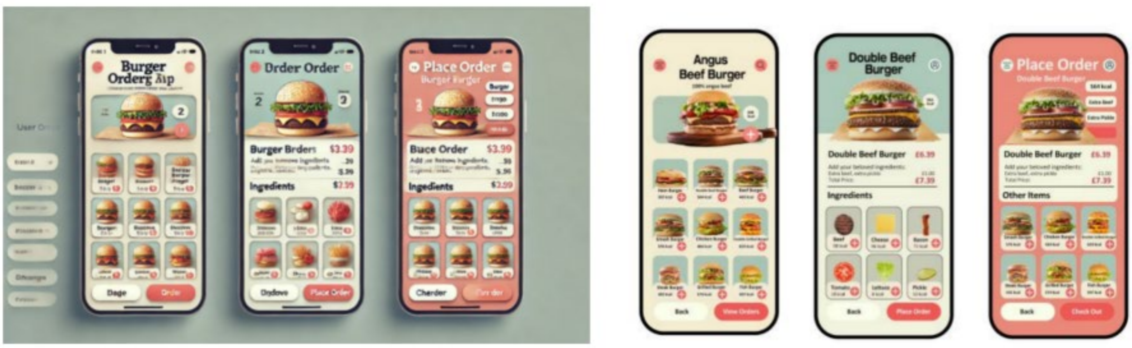


Figure 3: User Interfaces generated by DALL-E 3 (left) modified versions used for evaluation

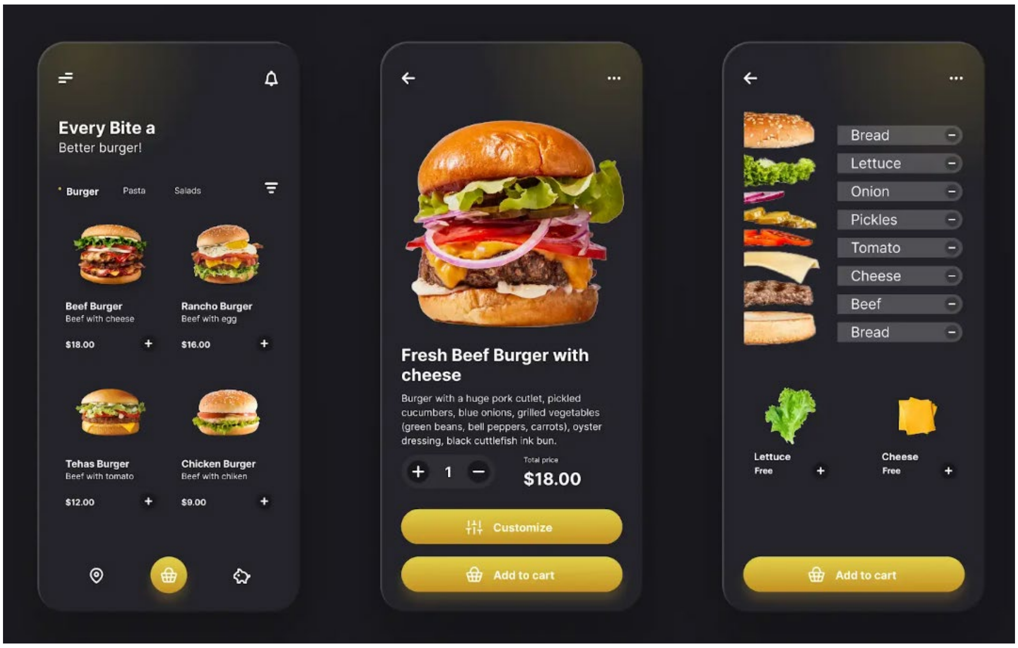


Figure 4: Example of HCIstudent designed user interface

The 20 sets of user interface designs were evaluated by 32 participants (recruited from students in the School of Computer Science at the University of Birmingham). Evaluation was performed using the User Experience Questionnaire-Short, UEQ-S, from Schrepp et al. (2017), a 7-point Likert scale for hedonic (i.e., look and feel), pragmatic (i.e., perceived ease of use) qualities.

Table 1: Items in UEQ-S

Negative	1	2	3	4	5	6	7	Positive
Obstructive								Supportive
Complicated								Easy
Inefficient								Efficient
Confusing								Clear
Boring								Exciting
Not Interesting								Interesting
Conventional								Inventive
Usual								Leading Edge

We opted for a rating scale, rather than textual answers, because we felt that this would provide a comparable set of data across the sets of user interface. We also note that the evaluation is performed on static images. As such, we are not reporting usability testing of a functional prototype. We were not sure how to render the user interface images dynamically for the GenAI to actively evaluate these and as such kept the evaluation consistent between human and AI rating activities.

Each set images of the user interfaces with a copy of the UEQ-S rating scale was presented on a page on GoogleForms. Respondents completed all 20 forms to evaluate the sets of user interface.

Prior to statistical analysis, we grouped the responses for the 5 user interface designs from each designer (HCI students, market apps, Dall-E 3, Stable Diffusion) and took the median rating from each participant for each designer. While the resulting data followed a normal distribution (according to Shapiro Wilke testing), it did not satisfy homogeneity and so a non-parametric, one-way Analysis of Variance was applied to the data for the designers.

Results

Human rating of user interface designs

In this section, we focus on the ratings for Pragmatic and Hedonic aspects of the user interfaces. These scores are calculated from combinations of the individual elements in the rating scale (table 1) following the guidance from the UEQ-S website[3].

There was a significant main effect of designer on Pragmatic scores [$x^2(3) = 19.3$, $p<0.005$]. Post-hoc, pairwise comparison (using Wilcoxon tests) revealed significant difference ($p<0.005$) between the designs from HCIstudents (median rating of 20) and the market apps (median rating of 16.5) and significant difference ($p<0.005$) between HCIstudents and Stable Diffusion (median rating 17). A possible explanation for the difference lower rating of the market apps is that had been modified (by removing identifying features prior to the evaluation) which might have affected their appearance. A second explanation might be that the designs from the HCI students and genAI followed the same set of requirements, and the market apps may have had different design intentions. The difference between the designs created by HCI students and Stable Diffusion suggests that the genAI were perceived less favourably. However, the rating of designs from Dall-E (median rating 20) was comparable to those of the HCI students. This suggests that Dall-E 3 was capable of creating a user interface that was comparable to one designed by novice HCI designers.

For Hedonic rating, there was a main effect of designer [$x^2(3) = 29.8$, $p<0.0001$]. The market apps (median rating 11) were significantly different ($p<0.0001$) to those designed by HCI students

[3] https://www.ueq-online.org

(median rating 16.5), also ($p<0.0001$) to those generated by Dall-E 3 (median rating 18), and also ($p<0.001$) Stable Diffusion (median rating 16). There were no differences between the genAI designs and HCI students' designs, although there was a significant difference ($p<0.05$) between ratings for Stable Diffusion and Dall-E 3.

In summary, Dall-E 3 produce user interface designs that were comparable to those created by novice interface designers. Given the modifications made to the commercial app designs, we cannot confidently conclude whether Dall-E 3 could produce professional-level user interface designs.

Using genAI to perform user experience evaluation

We designed prompts to apply UEQ-S by ChatGPT4o and Stable Assistant and presented the 20 user interface designs to these for evaluation. We found that the ratings on UEQ-S were more consistent for the gen-AI tools than the human participants. However, the gen-AI tools tended to use a restricted set of the rating scale rather than the full range. There was little correlation between human and gen-AI ratings. We scaled the human ratings (i.e., collapsing ratings of 1, 2 as 'low', 3,4 as 'neutral' and 5, 6, 7 as 'high') to reflect the 3 points used by gen-AI ratings but this did not improve correlation. It seemed that gen-AI was not applying the rating scale in the same way as people were. We tried changing the prompts but this did not improve matters.

Conclusions

GenAI, with appropriate prompt engineering and careful screening, is able to generate user interface concept designs that were comparable to those produced by competent designers. Given the caveats relating to the modified versions of the user interfaces from the commercial apps, we are not able to comment on how the designs relate to those from expert designers. However, at present, GenAI struggles to perform user experience evaluation to a level that compares with humans. This could be due to prompting or to a lack of knowledge or experience in interacting with user interfaces. If the problem relates to the latter point, then modifying GenAI to acquire such experience could be a next step to explore.

References

Cropley, D., 2023, Is Artificial Intelligence More Creative Than Humans?: ChatGPT And The Divergent Association Task. *Learning Letters*, *2*, 13.

Koivisto, M. And Grassini, S., 2023, Best Humans Still Outperform Artificial Intelligence In A Creative Divergent Thinking Task. *Scientific Reports*, *13*, 13601.

Lee, H.K., 2022, Rethinking Creativity: Creative Industries, AI And Everyday Creativity. *Media, Culture & Society*, *44*, 601-612.

Aliyev, E., 2023, Risk Assessment Of Using Artificial Intelligence Systems In Creative Human Activities. *Reliability: Theory & Applications*, *18*, 238-250.

Takaffoli, M., Li, S. And Mäkelä, V., 2024, Generative AI In User Experience Design And Research: How Do UX Practitioners, Teams, And Companies Use Genai In Industry?. In *Proceedings Of The 2024 ACM Designing Interactive Systems Conference*, 1579-1593).

Pourasad, A.E. And Maalej, W., 2024, Does Genai Make Usability Testing Obsolete?, *Arxiv Preprint Arxiv:2411.00634*.

Ronanki, K., Cabrero-Daniel, B., Horkoff, J. and Berger, C., 2024, Requirements Engineering Using Generative Ai: Prompts And Prompting Patterns, *Generative AI For Effective Software Development*, Cham: Springer Nature Switzerland, 109-127.

Schrepp, M., Hinderks, A. and Thomaschewski, J., 2017, Design and evaluation of a short version of the user experience questionnaire (UEQ-S, *International Journal of Interactive Multimedia and Artificial Intelligence, 4*, 103-108.

Human-AI Decision-Making in Higher Education

Robert Houghton, Xinrui Zhai, Zhuojun Li & David R. Large

Human Factors Research Group, University of Nottingham, UK

SUMMARY

Findings from an online study and focus groups show that students and staff were receptive to AI decision-making in higher education, highlighting transparency and equitability. Even so, students expressed a desire for creative work to be seen by a 'fellow human', suggesting potential for a collaborative human/AI approach. However, a follow-up study revealed that this solution was in fact perceived as less desirable than either a human- or, in some cases, an AI-only decision maker.

KEYWORDS

Collaborative Human-AI Decision Making, Higher Education, User Perception

Introduction

In Higher Education (HE), Artificial Intelligence (AI) has the potential to enrich and extend learning outcomes and experiences, and to reduce the administrative workload for educators, by optimising teaching resources and providing algorithm-driven marking and assessment (Zawacki-Richter et al., 2019). However, the emergence of AI is an emotive and polarising topic, and its adoption may be perceived as de-valuing or displacing humans. Moreover, education presents a unique context in which trust can be fragile, and institutions must therefore ensure clarity and transparency in decision-making. In practice, *perceptions* of AI decision-making, regardless of the actual quality of the decision outcome, can significantly influence confidence and trust, and result in poor acceptance or rejection of the system (Sundar & Nass, 2001). AI has traditionally been associated with big data, statistics and machine learning, and this enables efficient, optimised, and data-driven decision-making. AI is also perceived as more rational and less emotional than people (Waytz and Norton, 2014). In a HE context, AI decision-making may thus be more suited to tasks that require 'mechanical' or algorithmic skills (e.g. timetabling and admissions, in which large datasets are managed and optimised). In contrast, tasks in which the decision maker is required to make subjective and intuitive judgements, or to understand and express emotions or navigate difficult social situations (e.g. marking written or creative work or resolving interpersonal conflicts), may be more suited to a human decision-maker (Lee, 2018). In practice, this distinction may be somewhat blurred, with many tasks requiring elements of both skillsets. It has therefore been suggested that a potential solution is to harness the complementary strengths of human and AI within collaborative decision making, in which AI does the 'heavy (data) lifting' and a human expert subsequently checks or refines the outcome (Dolgikh & Mulesa, 2021). Indeed, *performance* successes have been reported with such an approach, but it is unclear how this is *perceived* by students and educators, which could ultimately affect adoption.

Method

In the first investigation, we explored attitudes towards human-only and AI-only decision-making in an HE context using both a survey (n=94, comprising students and academic staff) and two focus groups (each comprising 4 survey respondents) to determine whether attitudes differed based on the nature of the task (i.e. tasks requiring mechanical compared to human skills). Building on the findings from the first survey, we subsequently conducted a second survey (n=75), in which we maintained the same task distinctions and academic setting but offered a third option, notably, a human-AI collaboration. In both surveys, which were hosted on https://www.prolific.com, participants were presented with four scenarios: admissions (academic and holistic assessment of candidate), marking (multiple-choice exam), course scheduling, marking (written essay). Inspired by Lee (2019), these differed in the degree to which they involved 'mechanical' and 'human' skills. In the first study, participants were told that for each scenario, the decision outcome was provided by either an experienced/expert human professional (e.g. administrator, professor, admissions officer, as appropriate) or an AI system, in isolation. In study two, a third option was added, in which the decision was made by AI and then "tweaked", "reviewed" or "adjusted" by the relevant human expert (Human-AI collaboration). Participants were asked to rate the fairness/accuracy, trustworthiness and emotional response/satisfaction associated with each scenario/decision maker, using 7-point Likert scales, where 1 indicated the least positive rating (typically, 'not at all'), and 7, the most positive (typically, 'completely', although the precise nomenclature and scale anchors depended on the factor under evaluation). Ratings were compared using one-way ANOVAs.

Results

Results from study one show that AI decision-makers were perceived as fairer and more trustworthy than human decision-makers in tasks requiring ***mechanical skills***; emotional response was similarly high for both AI and humans. Higher ratings of fairness and trust were attributed to the efficiency and objectivity of the AI system when executing mechanical tasks, in which the decision-making process relied upon analysing facts and data against predefined rules and algorithms. In contrast, human decision-makers were considered to be susceptible to subjective emotions and personal preferences that could affect outcomes. Nevertheless, while participants stated that AI offered fairness and was "immune to exhibiting discrimination", it was also highlighted that human decision-makers offered "more profound insight" and could adapt their decision-making to take into account unexpected human factors that might only come to light when a decision is implemented. The positive emotional response associated with AI decision-making reflected the belief that AI exhibited "consistency, efficiency and impartiality" and its decisions were "transparent". Surprisingly, in tasks requiring ***human skills***, participants also perceived AI decision-makers as fairer than human decision-makers, and also equally trustworthy/reliable, although the emotional response towards AI was less positive for these tasks. This outcome reflected the potential for unfair or biased decision-making in situations where humans made the decision, for example, the "potential for nepotism when deciding on admissions". However, it was also recognised that human decision-makers have a sense of responsibility and accountability, especially when problems or mistakes occur, and that the same accountability could not be attributed to AI as it lacked agency. Notably, this did not preclude the concern of algorithmic bias during AI decision-making, or indeed, any technical limitations (for example, in making 'value' judgments, see: Araujo et al., 2020), but it was felt that if any issues arose with an AI decision, it would be "transparent" and could be reviewed and corrected in a fair and impartial manner. As such, participants were generally supportive of algorithmic marking or qualitative AI assessment, for example. However, there was a notable concern expressed that students would feel "profoundly disrespected" if their written assignment or application was evaluated solely by AI and was never seen by a "fellow human". This supports the notion of a hybrid approach, in which an AI decision could be checked or reviewed by a human.

However, results from study two suggest that ***collaborative Human-AI decision-making*** was generally considered to be less accurate, less trustworthy and likely to provide lower satisfaction in the decision outcome, compared to a human decision-maker, and in some cases, also to AI in isolation. This is perhaps surprising given the recommendations and expectations from study one, and, indeed, the performance benefits this hybrid approach purportedly offers (Dolgikh & Mulesa, 2021) and suggests that in an HE context, at least, a hybrid approach may fall short of its 'best-of-both' ambition.

Results may reflect a lack of understanding of precisely how human and AI could work collaboratively in a HE context, but also reinforce the importance of determining how the decision-making process is perceived, not only how well it performs, particularly by those people who are directly impacted by the outcome.

Acknowledgements

The study received funding from the Engineering and Physical Sciences Research Council (EPSRC) as part of the 'Made Smarter Innovation - People-Led Digitalisation' project (EP/V062042/1)

References

Araujo, T., Helberger, N., Kruikemeier, S. and De Vreese, C.H., 2020. In AI we trust? Perceptions about automated decision-making by artificial intelligence. AI & society, 35(3), pp.611-623.
Dolgikh, S. and Mulesa, O., 2021, September. Collaborative Human-AI Decision-Making Systems. In IntSol Workshops (pp. 96-105).
Lee, M.K., 2018. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. Big Data & Society, 5(1), p.2053951718756684.
Sundar, S.S. and Nass, C., 2001. Conceptualizing sources in online news. Journal of communication, 51(1), pp.52-72.
Waytz, A. and Norton, M.I., 2014. Botsourcing and outsourcing: Robot, British, Chinese, and German workers are for thinking—not feeling—jobs. Emotion, 14(2), p.434.
Zawacki-Richter, O., Marín, V.I., Bond, M. and Gouverneur, F., 2019. Systematic review of research on artificial intelligence applications in higher education–where are the educators? International Journal of Educational Technology in Higher Education, 16(1), pp.1-27.

Improving ineffective instructions

Andy Brazier

AB Risk Ltd

SUMMARY

Despite the availability of extensive guidance, many instructions, procedures, and other documents intended to control how work is performed fail to fulfil their primary role. Major incidents and day-to-day operational inefficiencies often reveal that written instructions are either misunderstood, ignored, or incorrectly followed. Style guides typically focus on aspects such as terminology, tense, and reading level, aiming to ensure clarity and readability. However, these guides rarely address the more critical question: what content should be included, or excluded, to make instructions truly effective.

Effective instructions go beyond simply describing how a task should be performed; they are tools that support users in completing tasks with greater accuracy, reducing errors, and enhancing overall consistency. The distinction between well-crafted instructions and poorly designed ones is not trivial; it can mean the difference between safe, efficient operations and incidents with serious consequences. This paper explores why many instructions fall short and how adopting a user-centred, task-focused approach can lead to better outcomes for individuals and organisations alike

KEYWORDS

Instructions, procedures, risk, safety, good practice, user centred

What is an effective instruction?

An instruction is a means of guiding someone to perform a task. When written, it defines the method to be followed. Other terms such as procedure, method statement, or guide are often used interchangeably. This paper primarily focuses on written instructions used in the workplace.

An effective instruction helps individuals perform a task correctly. In situations where safety is a concern, the chosen method should aim to reduce risks to As Low As Reasonably Practicable (ALARP). Even when several methods of performing a task are equally safe and efficient, there can be benefits in ensuring that everyone performs the task in the same way.

To be effective, instructions must be technically correct and up to date. However, that alone does not guarantee they will be used or followed correctly. People must be aware that the instructions exist and be able to locate them easily. Also, they must be willing to use the instructions as intended.

Everyday experience of instructions

Instructions are prevalent in everyday life. Supplied with items we buy and presented when using devices. Whether anyone actually reads them is another matter. We might skim through the instruction booklet the first time we use a new washing machine, but after that, it is set aside and only consulted if something goes wrong. While we probably use only a fraction of the available wash cycles on our sophisticated machine, as long as our clothes are clean, we are satisfied.

With flat-pack furniture we are likely to follow the instructions very closely, knowing there is little margin for error. Yet, if we are assembling multiple items, such as a set of dining chairs, we will pay far less attention as we become more skilled with the process.

Evidence that (some) instructions in the workplace are ineffective

The UK Health and Safety Executive's document Revitalising Procedures [HSE 2004] identified poor procedures as a contributing factor in major accidents, including Bhopal (1984), Piper Alpha (1988), and the Clapham Junction rail crash (1988). It stated, "The consequences of inadequate procedures, or operators not following procedures, can be disastrous."

A recent review of letters sent by the HSE to offshore oil and gas installations revealed that similar issues persist [Salus 2023]. The review highlighted numerous instances where procedures and instructions were described as "insufficient" and not reflective of how tasks are actually performed in practice. Underlying problems included a lack of monitoring evidence, inadequate auditing, and failure to update procedures following reviews and operational changes.

Gregory Smith, in his book *Paper Safe* [Smith 2018], argues that instructions have become entrenched in bureaucracy, where paperwork has become disconnected from its core purpose of managing risks. This creates an illusion of well-managed safety that workers are content to play along with, and management is prepared to accept, because it has become normalised. Implementing improvements, however, would require significant change, which is often met with resistance.

Underlying issues that make written instructions ineffective

Ambiguous instructions leave too much room for interpretation. Overly complex language and long-winded explanations exacerbate this by making it harder for people to understand and follow the steps. Simpler, more direct wording, cutting any content that doesn't directly support the person performing the task, must be a key objective for authors.

The human brain has limited working memory, and instructions that overwhelm with excessive information cause overload and confusion, leading to skipped or mistimed steps. A well-structured instruction with a mechanism for users to track their progress can help overcome these natural human limitations.

Small font sizes, dense blocks of text, and messy formatting discourage people from reading instructions in detail. They may skim for key points and assume they know the rest, a tendency fuelled by natural human biases that promote overconfidence. A clean, visually pleasing layout increases the likelihood that the content will be read and followed. Minimising word count and presenting the instruction in a clear, structured format makes this far easier to achieve.

Time pressure and perceived priorities strongly influence how people use instructions. Significant cultural issues must be addressed to ensure these factors don't affect how critical tasks are performed, allowing risks to be managed to ALARP. A well-designed system for managing instructions supports the development of a suitable safety culture. Attempting to change culture without good systems is rarely successful.

Instructions that are misaligned with users' competence, either too advanced for novices or overly detailed for experienced personnel, affect how they are perceived and used in practice. Language proficiency and cultural differences can create additional barriers. Understanding the user population is essential, and a one-size-fits-all approach may not be appropriate.

A lack of feedback to users may lead them to assume they are following the instruction correctly. Similarly, a lack of feedback to authors may leave them believing their instructions are fit for purpose. A collaborative approach, including coaching for both users and authors, should be an integral part of the instruction management process.

Written instructions are inherently limited as a communication method for critical information. If users do not fully understand the task and associated risks, they are more likely to make assumptions and less equipped to handle unexpected situations. A semi-structured pre-task briefing can help ensure people understand the task they are about to perform. This briefing can be enhanced by written instructions that guide and support the topics to be discussed.

User-centred focus

Guidance often emphasises the importance of involving end users in the development of instructions. However, this can result in end users being tasked with writing the instructions themselves, often leading to ineffective outcomes. This is not the fault of those individuals, as they are typically neither trained for the task nor given adequate support.

End users may feel pressured to create idealised instructions, including every possible detail, to avoid criticism for omitting something important. While they may possess practical experience, their technical understanding is sometimes incomplete. One key purpose of instructions is to help users perform tasks in accordance with recognised good practices, which may require them to change current practices [Brazier, 2024].

Handing responsibility over to technical authors, who may identify as professional procedure writers, often results in well-presented documents. However, technical authors rarely have the technical expertise or hands-on experience required to create accurate and effective procedures. Although their involvement may improve KPI scores, the quality of the content can be lacking.

A collaborative, iterative approach should be adopted wherever possible. Perfection is unattainable, and attempting to achieve it can be counterproductive. The focus should remain on creating an instruction that is practical and genuinely useful to the user.

Name and number

Being able to find the correct instruction will reduce one of the barriers to use. Giving it a meaningful title using terminology the end users will understand will make sure they identify it easily. If there are similar procedures for different scenarios, indicating this in the title is better than in the main body because that requires the user to access the full procedure to decide which one they need to use.

A task numbering convention can be very useful. Using a unique number for each task and a code for different document types allows easy cross referencing and access to information. For example:

- OI and MI to indicate an operating or maintenance instruction;
- JA for job aids that may provide additional support (e.g. checklist, flowchart, diagram);
- RA for a risk assessment;
- SCTA for a safety critical task analysis.

The document number for an operating instruction for task 123, performed in area 01 of the NTH site may be NTH-01-123-OI, whilst the associated SCTA is NTH-01-123-SCTA.

Writing more effective instructions

While some instructions may be intended for use by the general population, most in an industrial setting are designed for individuals who possess a certain level of competence and either already know how to perform the task or have experience with similar tasks. The primary purpose of these instructions is to reduce the likelihood of errors and enhance task consistency. Including unnecessary detail diminishes an instruction’s effectiveness and may lead competent users to disregard it entirely. Those learning a task may require additional support, but this should be provided through other means, such as supervision or training guides, and should not justify overloading instructions with excessive detail.

Being clear about how an instruction is to be used is crucial. It should not be left to the discretion of the end user. For instance, if users are expected to print, follow, and sign a specific instruction each time it is used, this requirement must be explicitly stated. Conversely, it is acceptable for some instructions to serve only as guidance, as long as this is clearly communicated.

It is also important to acknowledge that following an instruction exactly as written may not always be possible or desirable. However, deviations must be managed through a clearly defined process. Furthermore, it should be emphasised that blindly following an instruction is not inherently safe.

The first pages of many instructions are often filled with background and generic information that holds little relevance for the end user. In most cases, a good title makes the instruction’s scope and purpose clear. However, when a standard template includes headings for these, authors often feel compelled to fill them with unnecessary content, increasing the overall word count and making it more likely that users will skip the entire preamble.

A Personal Protective Equipment (PPE) section is often mandated, which at first glance makes sense due to safety implications. However, in many cases, the requirements listed are not task-specific. If special PPE is required for a particular task, it is usually controlled by another mechanism, such as a Permit to Work.

However, a preamble that supports a pre-task briefing could significantly enhance the effectiveness of instructions by preventing people from diving straight into tasks without thinking and creating a forum for discussion and information sharing. Useful information may include:

- Main task stages;
- Major accident hazards associated with the task and associated risk controls;
- Parts of the task vulnerable to human error or where human actions are relied on to control the risk;
- Unique aspects that require different approach from other similar tasks.

The aim should be minimisation. Standard preamble sub-sections in a template often encourage authors to add content, even if it provides little value. Technology is now offering excellent opportunities to support preparation. For example, 3D-scanned images of work areas and smart drawings allow people to virtually walk through a task from a safe and comfortable location.

Warnings, cautions and supporting information

Current guidance for writing instructions often emphasises the use of warnings and cautions as a key safety feature, specifying that they should appear before the associated step, since reading a warning after encountering the hazard is too late. However, implementation is often poor. In some instructions, warnings and cautions outnumber the actual task steps, with the provided information frequently duplicating or even contradicting the steps that follow. In practice, warnings and cautions are rarely as important or useful as they seem.

More fundamentally, including warnings and cautions can suggest that some steps are more important than others. If certain steps are unimportant, why include them at all? It can also imply that steps without warnings are optional. Covering hazards and controls during a pre-task briefing, supported by the instruction preamble, is a far better way to ensure that workers understand the risks they will encounter. The only situation where a warning or caution might be advisable is when a task step relies entirely on human vigilance to avoid a significant consequence, something we usually try to eliminate. When such steps are identified, it is crucial to review whether the hierarchy of risk controls has been applied correctly.

While warnings and cautions have questionable value, a mechanism for providing additional information for complex or critical steps can be highly beneficial. It allows task step descriptions to remain concise and to the point, while the supporting information is especially helpful for less experienced individuals. A second column alongside the task steps is a simple and effective way to achieve this. Not every step will require additional details, and the text should be kept as brief as possible.

Diagrams, photos, and other visual aids can also provide valuable support and are encouraged by existing guidance. However, practical implementation is often poor, and the significant effort required to embed visuals directly into instructions is rarely justified. It is usually more efficient to include them as appendices or capture them as separate job aids linked by the task number.

Keeping place

When skipping or duplicating steps during a task can have serious consequences, it is essential to provide users with a reliable way to track their progress. This becomes especially important if the task extends beyond a shift handover or involves a change in responsibility.

A misplaced focus on accountability has often led to instructions requiring every step to be signed and dated upon completion. This process can be laborious and may encourage users to sign off all steps at the end, defeating the purpose of keeping track. A simple tick box for each step is usually sufficient. However, users may sometimes be reluctant to rely on tick boxes, fearing that someone else could mark them on their behalf. This may reflect deeper cultural issues that warrant investigation. To address this concern, providing a box next to each step and allowing users to mark it with a tick, signature, or other symbol can meet the primary goal of place keeping.

Critical hold points should be clearly identified and provide a good opportunity to record date, time and who was involved at different stages of a task. At a minimum, a hold point should appear after confirming preconditions at the start of the task and at the end task. The concept of 'postconditions' from computer programming could be useful for critical tasks to define the system state on completion.

Open-ended steps such as monitoring a condition while other steps are performed or contingent steps that cannot be predicted with precision present a subtle challenge for place keeping. In these cases, the requirement should be for the user to acknowledge awareness of the action rather than confirm its immediate completion. A tick in the box can indicate that acknowledgment rather than execution. A subsequent critical hold point may provide a useful opportunity to confirm that the monitored condition has remained within limits and whether any contingent step was performed.

Whenever possible, tasks should be completed by the same person or team. If that is not feasible, the goal should be to hand over responsibility at a critical hold point defined in the instruction. If these are impractical a method for indicating the transfer of responsibility should be provided. Drawing a line and noting the change of responsibility is often sufficient.

Effective text

Guidance consistently emphasises that the wording of instructions should be clear and concise. However, these are subjective qualities that authors often find difficult to achieve in practice.

Wording is not trivial and should be subject to greater scrutiny than is typically the case. Formal task analysis is a useful tool for this but it is not practical to analyse every task requiring an instruction.

Accurately identifying items of plant and equipment is critical. Redundancy in information helps reduce the likelihood of confusion. For example, using both a description and a tag number for each item of equipment should be mandatory so that users have at least two ways to confirm identity of the correct item.

Another way to minimise confusion is to use consistent terminology throughout the instruction. It is common for the same valve to be referred to as both ‘pump suction’ and ‘pump inlet’ in different steps of the same instruction. A practical solution to address both issues is to list all items (e.g., valves, instruments, equipment) with their specific descriptions and tag numbers at the start of the instruction, ensuring consistency throughout.

Each line of the procedure should be reviewed to minimise word count while maintaining clarity. Standard guidance to use the active voice greatly improves readability. Simple changes, such as removing unnecessary words like ‘the’ and using symbols (e.g., ‘&’ instead of ‘and’), can also help. Although guidance often warns against using abbreviations and acronyms, they are perfectly acceptable if they are familiar to the instruction’s users.

Consistency and redundancy in equipment descriptions also offer opportunities for further word reduction. For example, the step "Ensure that the acid pump suction valve V101 is closed" can be simplified to "Ensure acid pump suction V101 is closed." Reducing the word count from ten to seven is useful, but more importantly, the focus it brings to the wording improves clarity and reduces cognitive load.

Using a hierarchical structure helps convey why a task must be performed in a particular way without needing to explain it. There are typically up to four levels of hierarchy as follows (task is to start oil export to pipeline):

1. Sub-task - Start export P101
    - 1.1 Sub-sub-task - Line-up export P101 discharge to pipeline
        - 1.1.1 Task step - Open discharge V1212
            - 1.1.1.1 Sub-step - Turn valve wheel anti-clockwise to full extent.

Each level of hierarchy in an instruction defines how the level above it is performed. The fourth level of hierarchy (sub-step) represents a level of detail that may be necessary for novices. This level of detail may be appropriate in everyday situations such as flatpack furniture instructions specifying that a screw should be turned clockwise to tighten. But in a workplace setting, it would be concerning if complete novices were expected to follow an instruction without some relevant skills or training.

It is important to recognise that technical perfection in an instruction is both unattainable and often counterproductive. Instructions are just one component of a broader system for controlling risk, and an imperfect instruction is far better than having no instruction at all.

Presentation

Current guidance on the presentation of instructions is generally sound. Using an appropriate font type and size, adequate line spacing, and ensuring plenty of white space on the page all help users read and comprehend the content more easily. However, consideration of neurodivergence may not have received enough attention. For example, highlighting keywords in capital letters or bold text can be very distracting for individuals with dyslexia. Similarly, the use of colour can be problematic for those with colour blindness.

There is no clear evidence that these emphasis techniques significantly improve comprehension for most users. Therefore, sticking to standard sentence structure and black text is a more sensible approach.

Conclusion

Current guidance on writing instructions has not succeeded in eliminating ineffective practices. While the guidance itself may not be fundamentally flawed, it often fails to focus on what end users actually need. The key to improving instruction lies in shifting the mindset from simply describing a task to genuinely supporting the people performing it. This change in focus significantly alters the content and structure of instructions, making them more practical and user-centric.

An effective instruction should prioritise usability, be concise, and avoid unnecessary detail. Being ruthless with wording, stripping away anything that does not directly aid task completion, results in clearer, more accessible instructions. It is also crucial to recognise that perfection is unattainable; striving for it often leads to bloated, overly complex instructions.

Thoroughly understanding the task and engaging in meaningful collaboration with end users allows authors to write instructions that align with real-world practices. This collaborative approach assists continuous improvement by bridging the gap between documented instructions and actual work practices.

Instructions should support users to perform tasks safely and consistently, not overwhelm them with unnecessary complexity or create a false sense of security through bureaucratic box-ticking. A user-centred approach that involves task analysis, ruthless editing, and practical collaboration is essential for improving the effectiveness of workplace instructions and reducing the risk of human error.

Although the terms have been used interchangeably in this paper, it may be concluded that there is a distinction worth recognising. The following definitions highlight different uses, which may affect both content and presentation:

- Instruction – A document created by one person or party to direct another on how to perform a task.
- Procedure – A document written to support a competent person in performing a task.
- Method Statement – A document prepared by one person or party to explain to another how a task will be performed.

Often, changing the style of an instruction to a procedure results in a simpler document that is more useful to end users with a defined level of competence. Instructions are better suited for novices, but whether we should be writing for novices in a hazardous workplace is debatable. The value of method statements, typically created by contractors on the insistence of the client is a topic beyond the scope of this paper.

Summary table

The table is a convenient summary of issues observed with written instructions (and procedures), and how they may be resolved.

Problem	Solutions
Not aware instruction exists	Make sure they only developed when necessary / useful. Delete / archive others. Ensure good communication with end users for new / changed.
Cannot identify correct instruction	Use clear / unambiguous titles. Be consistent with end users' terminology.
Cannot find instruction	Organise and index following a consistent structure. Use multiple methods of identification (e.g. description and number).
Cannot access instruction	Understand how end users access in practice. Ensure end users have the necessary permission to access. Provide controlled hard copies where necessary.
Not sure of correct version	Apply robust version control. Make sure superseded versions are removed / archived. Make sure the correct method of access is efficient / discourage local copies.
No instruction for task	Focus resources where necessary / useful. Develop at the earliest possible stage of projects and task planning. Acknowledge not required for every task.
Out of date instructions	Make sure updated as part of management of change. Make sure reviews are focused on how tasks are performed. Ensure review and authorisation of new / changes occurs promptly.
Instruction not consistent with way task performed	Modify plant / equipment so that the method preferred by end users is the safest / most efficient. Involve end users in development / review. Ensure feedback identifies when updates needed. Be aware that end users may follow unsafe methods and preferred method needs to be better communicated / enforced.
Instruction does not describe safe method	Involve technical experts in development / review Use task analysis to identify safe and practical methods. Cross reference safety studies.
Instruction not used as intended / expected	Clearly define intended use. Make sure expectations are realistic for end users. Implement a deviations process to cater for real world scenarios. Accept that blindly following instructions is not safe. Develop safe compliance culture.
Instruction misinterpreted	Be ruthless with wording. Use hierarchical structure to convey understanding. Make sure presentation is accessible for neurodivergence (autism, dyslexia, colour blindness). Accept procedures always open to interpretation and must have flexibility to cover uncertainty.

Problem	Solutions
Key messages overlooked	Only include content that helps people understand. Minimise use of warnings / cautions. Accept written communication is relatively poor at conveying meaning.

References

Brazier, Andrew. (2024), Human factors role in supporting best practice.

HSE. (2004), Revitalising Procedures.

Salus Technical. (2023), How offshore inspection scores reveal major accident prevention measures.

Smith, Gregory. (2018), Paper Safe: The triumph of bureaucracy in safety management. Wayland Legal. Kindle Edition.

Operational Strategies within Simulated Environments

Mohsen Zare[1], Bernard Mignot[1], Nicolas Bert[1] & Maxime Norval[2]

[1]UTBM, [2]Université Orlean, France

SUMMARY

Exploring the range of operational strategies could enhance our understanding of effective regulatory approaches, contributing to the prevention of Musculoskeletal Disorders (MSDs). This study aims to assess whether the use of virtual simulation, integrated into a Serious Game, enables users to establish conditions conducive to efficient regulation of their future professional activities. Additionally, we seek to demonstrate how operational strategies differ when performing similar tasks within the simulated context of a virtual reality Serious Game. The key findings reveal the presence of multiple operational strategies for each simulated task, which appear to establish favorable conditions for better regulation of future activities.

While the Serious Game demonstrated a diversity of strategies, an important question remains regarding whether the game led to improved postural behavior between the beginning and the end of the session, and whether these changes transferred to real work conditions. The analysis showed that certain users progressively adjusted their strategies based on in-game feedback, improving postures related to upper limb and back positioning. However, others maintained their initial approaches despite recommendations. The extent to which this learning translates into long-term workplace adjustments requires further study.

KEYWORDS

Virtual Reality, Operational Strategy, Serious Game, MSD Prevention

Introduction

Numerous automotive industries have attempted to train their operators in standardised best practices for performing assembly tasks, with the aim of mitigating health risks and addressing performance challenges. These efforts, however, have often fallen short of achieving their intended goals, as no universal best practices or standardised operational strategies exist for specific tasks. Instead, each operator tends to develop individualised strategies based on the resources and constraints available to them. Consequently, effective training programs should empower operators to explore and practice a range of diverse operational strategies, enabling them to adapt more effectively to the varied situations encountered in the workplace.

Developing an innovative tool that facilitates the creation of context-specific operational strategies represents a promising approach to this challenge. Simulation-based learning, tailored to industrial applications, holds significant potential for enhancing operators' regulatory capacities. Digital technologies, such as virtual reality and Serious Games, offer a safe and controlled environment where operators can explore, experiment with, and refine their operational strategies.

The primary objective of this article is to propose an innovative approach to the prevention of MSDs through the development of a tool designed to help operators craft operational strategies that are specifically adapted to their work context. We aim to demonstrate how real industrial activities vary among operators and to explain how a virtual tool, integrating scenarios based on these real-world activities, provides users with the opportunity to develop and adapt their operational strategies in response to challenges encountered during gameplay.

Methods

The game environment was designed to be engaging and relatable to real-world tasks while maintaining originality to avoid a purely industrial feel. A tutorial introduced users to the virtual environment, where a realistic avatar mirrored their movements, enabling direct interaction with virtual objects. Players could choose between immersive environments (beach, forest, snowy mountain) that gradually transitioned into workplace settings. A virtual trainer provided personalised feedback on postural risks and performance after each level. Players accessed their scores and tailored recommendations after each trial for improvement and experimented with various operational strategies. Postural risks were evaluated using a color-coded scoring system for upper limb joints, time spent in risky postures, and task performance metrics (Figure 1).

Figure 1: color-coded scoring system for upper limb joints

Figure 2 illustrates key aspects of the virtual environment and gameplay scenarios, depicting how operators interact with the system and adjust their strategies.

Figure 2: Virtual environment and interaction with the game

Results

We performed quantitative and qualitative analysis of playing serious game scenarios in virtual reality to explore the diverse operational strategies employed by subjects during gameplay. For quantitative analysis A test experiment with 44 operators (28 men, 16 women) from the real industrial sectors revealed significant variability in posture scores (based on exposure duration) and performance scores (completion time) for identical tasks, reflecting distinct strategies. For instance, in the "puzzle assembly" level, one Operator had a posture score of 6.53, indicating risky postures for the back, neck, and elbows, while another Operator, with a lower score of 2.57, adopted less risky postures but took more time to complete the task. Operators adopted varying strategies even for the same task. Serious Games appeared to offer users flexibility to explore different approaches.

Qualitative analysis performed by observing all the activities of seven subjects and interviewing with them using Vézina's activity-centred model indicators (2001) confirmed this diversity. Notably, one user showed minimal strategy regulations, ignoring game feedback and maintaining the same workstation setup throughout. Conversely, another subject adapted strategy by adjusting workstation height and altering task execution based on in-game recommendations.

The game fostered progressive awareness among users regarding their operational strategies. This awareness, blending in-game learning and real-world experience, enabled users to better utilise workplace resources and adopt strategies tailored to health and performance demands. However, technical limitations of the virtual reality system, such as occasional lack of fluidity and restricted hand movement, sometimes hindered optimal strategy regulation.

A key question concerns whether playing the game improved postures and whether these changes transferred to the workplace. Our findings indicate that some players gradually refined their strategies, adjusting their movements and task execution based on feedback. However, technical limitations—such as restricted hand movement in VR—sometimes prevented optimal strategy regulation. Regarding real-world transferability, while the game effectively raised awareness, no direct observation of long-term workplace behavioural change was conducted in this study, necessitating further investigation.

Conclusion

The diversity of strategies observed in the Serious Game can be linked to internal factors like motor control systems, experience, and expertise, as well as external resources provided by the work environment or the game itself. However, further research is needed to understand how these factors influence strategy diversity and motor variability. The findings underscore the importance of tools that enable operators to explore diverse approaches to task execution and develop strategies suited to their characteristics, tasks, and professional demands. The Serious Game used in this study demonstrates its potential as a digital tool to simulate industrial tasks and foster strategies that support both health and performance.

The resistance of some operators to adjusting their strategies, despite game feedback, highlights an important challenge in workplace training. The study revealed that while the game fosters awareness, some operators did not alter their approach, suggesting the need for complementary interventions, such as ergonomic coaching or workplace adjustments, to facilitate real-world application of learned strategies.

Future developments could focus on virtual reality games that allow operators to adjust their work environment and organisation, enhancing the conditions for task execution. Our findings suggest that Serious Games encourage the development of diverse strategies, though this depends on users' familiarity with the game's features and prior experience. Further research is needed to explore how strategies learned in simulations can effectively translate to real-world applications.

References

Coutarel F, Caroly S, Vézina N, Daniellou F. Marge de manœuvre situationnelle et pouvoir d'agir : des concepts à l'intervention ergonomique. Le travail humain 2015;78:9–29. https://doi.org/10.3917/th.781.0009.

Coutarel F, Aublet-Cuvelier A, Caroly S, Vézina N, Roquelaure Y, Cuny-Guerrier A, et al. Marge de manœuvre et prévention des troubles musculo-squelettiques : quelles perspectives ? Le travail humain 2022;85:3–31. https://doi.org/10.3917/th.851.0003

Djaouti D. Serious Games pour l'éducation : utiliser, créer, faire créer ? Tréma 2016:51–64. https://doi.org/10.4000/trema.3386.

Mathiassen SE. Diversity and variation in biomechanical exposure: What is it, and why would we like to know? Applied Ergonomics 2006;37:419–27. https://doi.org/10.1016/j.apergo.2006.04.006.

Morais A, Aubineau R. Articulation entre l'ergonomie et le lean manufacturing chez PSA. Activités 2012;09. https://doi.org/10.4000/activites.468

Norval M. Les outils simples d'évaluation du risque d'apparition des troubles musculo squelettiques (TMS) : quelle intégration de la marge de manœuvre situationnelle (MMS) dans le cadre du repérage des situations à risques ? : étude de cas dans une industrie d'assemblage de moteurs diesel à usage non routier Operational. These de doctorat. Angers, 2019.

Sisto M, Zare M, Ouerhani N, Bolinhas C, Divernois M, Mignot B, et al. Virtual Reality Serious Game for Musculoskeletal Disorder Prevention, Cham: Springer International Publishing; 2018, p. 43–59.

Stergiou N, Decker LM. Human movement variability, nonlinear dynamics, and pathology: Is there a connection? Human Movement Science 2011;30:869–88. https://doi.org/10.1016/j.humov.2011.06.002.

Vézina N. La pratique de l'ergonomie face aux TMS: ouverture à l'interdisciplinarité. Comptes Rendus Du Congrès SELF-ACE 2001.

People Oriented Smart Towns

Barry Peter Kirby, Amanda Georgina Kirby & Jim Wilson

K Sharp Ltd

SUMMARY

The growth of interest in Smart Cities and development into Smart Towns and Smart Communities, has been driven by the technology development, however much of the data developed is not being utilised because it is not clear on its value. The People Oriented Smart Towns (POST) methodology has been developed to put user requirements at the front of Smart developments. This paper describes how that process is being applied to Kidwelly in South Wales and describes the outcomes so far.

KEYWORDS

Human Factors, User Centred Approach, Smart Towns, Smart Cities, IOT

Project Overview

Kidwelly, a small town with a rich historical background, has been actively seeking to capitalise on digital solutions to boost its local economy, streamline public services, and create a more inclusive and connected environment. The application of the POST (People-Oriented Smart Towns) methodology in the development of a Smart Town for Kidwelly has been an innovative effort aimed at transforming the town's infrastructure through digital technology while keeping the needs of the community at its heart. Unlike traditional Smart Town initiatives that often prioritise technology, POST shifts the focus to the people - residents, businesses, and visitors - ensuring that the technological enhancements are aligned with improving their day-to-day experiences.

The POST methodology applied to Kidwelly consists of two main phases: the Research Phase and the Action Phase, each of which is designed to ensure a deep understanding of the town and a structured approach to implementing technology that solves real-world problems.

Research Phase: The Research Phase is crucial to understanding the unique dynamics of Kidwelly, its population, and its infrastructure. It is structured around the following key activities:

1. ***Stakeholder Consultation***: The project began by consulting with a wide range of stakeholders. Initial meetings helped establish the project's scope, define geographical boundaries, and set goals tailored to the needs of each group.
2. ***People-Focused Discovery***: The project team segmented the local population into four key groups: residents, visitors, local businesses, and external stakeholders such as tourists and commuters. Through surveys, workshops, and informal discussions, data was gathered to create a detailed profile of how these groups interact with the town's services and infrastructure.
3. ***Systematic Investigation***: A comprehensive review of the town's existing digital infrastructure was conducted. This involved mapping mobile network coverage, analysing the use of existing public Wi-Fi services, and understanding how current Internet of Things (IoT) technologies, such as the LoRaWAN sensors, were being utilised.

4. ***Urban Dynamics***: This stage involved building a detailed picture of how people move through and interact with the town. The team examined the main routes taken by visitors and identified patterns of activity that could inform where future digital interventions would be most impactful.

Action Phase: The Action Phase focuses on implementing the insights from the Research Phase. It follows a cycle of planning, deployment, monitoring, and review to ensure the interventions meet the community's evolving needs.

1. ***Gaps and Remedies***: Based on the systematic investigation, several gaps in Kidwelly's infrastructure were identified, such as poor mobile connectivity, limited Wi-Fi coverage, and underutilisation of existing digital systems.
2. ***SMART Plan***: A clear implementation plan was developed, focusing on realistic, achievable milestones that involved all stakeholders.
3. ***Implementation and Verification***: In this step, the technological solutions are deployed and tested.
4. ***Monitor, Assess, and Evaluate***: A critical aspect of the POST methodology is its iterative nature, ensuring that the technology adapts to changing community needs.

Main Findings

Through the application of the POST methodology in Kidwelly, several important findings have been identified:

1. ***Low Awareness of Smart Town Initiatives***: Over 60% of residents and visitors in Kidwelly were unaware of the town's Smart technologies, with only 3% regularly using the free Wi-Fi.
2. ***Poor Mobile Connectivity***: The vast majority of respondents rated mobile connectivity in Kidwelly as poor, particularly indoors.
3. ***Demand for Real-Time Information***: Both residents and visitors expressed a clear desire for real-time updates on local transport, traffic, parking, and events.
4. ***Concerns Over Data Privacy***: Data privacy emerged as a significant concern, with over half of the respondents (54%) expressing discomfort about how their data might be collected and used.
5. ***Infrastructure Improvements Needed***: While Kidwelly has some foundational digital infrastructure, such as free public Wi-Fi and limited LoRaWAN sensors, these are currently underutilised.

Future Recommendations

Recommendations for Kidwelly	*Recommendations for the POST Methodology*
1.Public Awareness Campaign	1.Focus on Early Community Engagement
2.Enhance Digital Infrastructure	2.Tailoring Solutions to Specific Towns
3.Develop a Kidwelly Mobile App	3.Continuous Feedback Loops
4.Data Privacy and Transparency	4.Enhanced Focus on Cybersecurity
5.Data-Driven Decision Making.	

Conclusion

The application of the POST methodology in Kidwelly has highlighted the importance of a people-first approach to Smart Town development. By focusing on the real needs of residents, businesses, and visitors, the project has uncovered several key areas for improvement, from expanding digital infrastructure to addressing data privacy concerns. Moving forward, continued engagement, infrastructure upgrades, and data-driven decision-making will be critical to realising the full potential of a Smart Kidwelly. The POST methodology, with its iterative, flexible approach, provides a valuable framework that can be adapted to other towns seeking to improve their communities through smart technology.

Reimagining Rasmussen's risk management framework: a contemporary view on risk

Paul Salmon & Gemma Read

University of the Sunshine Coast, Australia

SUMMARY

Rasmussen's Risk Management Framework (RMF; Rasmussen, 1997) is arguably the most popular model of safety and risk within Ergonomics and Human Factors (EHF). However, the RMF was developed almost thirty years ago and there are questions regarding its suitability for contemporary systems and problems. In this presentation we outline and demonstrate a revised and extended RMF for contemporary sociotechnical systems.

KEYWORDS

Safety, Risk, Systems thinking

Introduction

Systems thinking is a core philosophy within Ergonomics and Human Factors (EHF) that is used to help understand and optimise performance and safety within complex sociotechnical systems. The philosophy is characterised by several models and methods which assert that safety and accidents are emergent properties arising from non-linear interactions between multiple components across work and societal systems (Leveson, 2004; Rasmussen, 1997). Rasmussen's risk management framework (RMF; Rasmussen, 1997) is arguably the most popular systems thinking model within EHF and safety science (Salmon et al., 2020) and along with the associated Accident Mapping (AcciMap) method has been used extensively over the past two decades. These applications have sought to understand and respond to complex problems in a diverse set of contexts ranging from the traditional EHF domains of defence, process control, healthcare, and transport to emerging application areas such as elite sport, outdoor recreation, and food safety (Salmon et al., 2020). Though the RMF's core propositions have been validated in these areas, flaws are evident, and the changing nature of work, technologies, societal issues, and global risks is beginning to expose gaps.

The aim of this paper is to return once again to the question posed in Rasmussen's seminal paper: *do we actually have adequate models of accident causation in the present dynamic society?* While the RMF has high utility and has provided important insights about safety and accident causation, given the changing context we argue that the answer is no. Therefore, we propose a revised and extended version of the RMF that is more suited to contemporary systems and problems. The revised RMF will be outlined in the conference presentation and demonstrated via a contemporary case study focused on the use of artificial intelligence in the higher education sector.

The revised risk management framework

Our work applying the RMF, Accimap, and associated models such as the Systems Theoretic Accident Model and Processes (STAMP; Leveson, 2004) in close to 20 domains has revealed important limitations. First, the current representation of complex systems as a hierarchy is artificial and overlooks dynamic interactions between actors that bypass 'system levels'. The focus on work at the bottom of the hierarchy also overlooks the fact that work occurs throughout the entire system. Second, though Rasmussen (1997) acknowledged the rapid pace of technological change, the RMF does not account for increasingly advanced non-human actors such as artificial intelligence technologies. This issue is compounded by rapid progress in the area and the likely future development of artificial general intelligence technologies that could surpass human levels of intelligence (Salmon et al., 2023). Third, central concepts of vertical integration and migration remain high level and have not been further developed, validated, or expanded, despite their critical role in system functioning. For example, how such concepts are influenced by contemporary societal issues such as misinformation, social media, artificial intelligence, pandemics, geopolitical tensions, and geostrategic shifts requires clarification.

Based on our extensive applications of the RMF, AcciMap, and STAMP, and the limitations described above, we propose a dynamic network-based RMF. This includes the following key features:

1. **Complex systems as a network**. Rather than viewing complex systems as a hierarchy, we instead propose that they comprise a dynamic network of actors and activities that are connected via control and feedback mechanisms. The number, relative power, and connectedness of actors differs across domains and is dynamic and changeable across situations.
2. **Human and non-human actors**. Complex systems comprise both human and non-human actors, with the latter becoming increasingly intelligent, autonomous, and influential. Rather than existing only on the frontline, intelligent non-human actors such as artificial intelligence are present throughout the system network and are increasingly enacting controls and providing feedback associated with supervisory, managerial and regulatory functions.
3. **Societal influences and pressures on behaviour**. An expanding range of interacting societal issues are having an increasing influence on behaviour within complex systems. Examples issues include the COVID-19 pandemic, economic and cost-of-living crises, societal health, misinformation, and current geostrategic shifts.

Discussion

Rasmussen's seminal RMF has had a significant influence on the discipline of EHF and continues to shape system safety research and practice in a diverse set of domains. In this presentation, we contend that refinement is required to support its continued use and relevance.

References

Leveson, N. (2004). A new accident model for engineering safer systems. *Safety science*, *42*(4), 237-270.

Rasmussen, J. (1997). Risk management in a dynamic society: a modelling problem. *Safety science*, *27*(2-3), 183-213.

Salmon, P. M., Hulme, A., Walker, G. H., Waterson, P., Berber, E., & Stanton, N. A. (2020). The big picture on accident causation: A review, synthesis and meta-analysis of AcciMap studies. *Safety science*, *126*, 104650.

Salmon, P. M., Baber, C., Burns, C., Carden, T., Cooke, N., Cummings, M., ... & Stanton, N. A. (2023). Managing the risks of artificial general intelligence: A human factors and ergonomics perspective. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *33*(5), 366-378.

The RAMP Package 2.0 for Sustainable Musculoskeletal Disorder Risk Management

Linda M Rose & Mikael Forsman

KTH Royal Institute of Technology, Sweden

SUMMARY

This paper describes the background to, and the development of the 'RAMP 2.0 Package'. This includes the enhanced application range in RAMP 2.0 with its Hand Model. The new web-based version of the RAMP 2.0 tool is presented together with a user-, and a reliability evaluation.

KEYWORDS

risk assessment, sustainable jobs, evaluation

Introduction

Musculoskeletal Disorders (MSDs) are still a large burden at societal, organisational and individual (EU-OSHA, 2019; Sobhani et al., 2015; Rose et al., 2014). In response to calls from industry, a systematic, research-based risk assessment and risk management tool for managing MSD risks was developed with the aim to support companies in developing sustainable jobs (UN, 2018). It was led by researchers in close collaboration with potential users of the tool.

In 2017 the result, the RAMP tool (Risk Assessment and Management tool for Manual handling proactively) was launched from a university website. In addition to the tool and the website, three Massive Open Online Courses (MOOCs) were developed and launched in 2018. These three parts formed the RAMP Package, available free of charge, which spread to over 100 countries rather rapidly. The RAMP tool was developed to align with the ISO standard 31000:2009 about the risk management process and addressed more relevant MSD risk factors than any other risk assessment tool (Rose et al., 2020). Although the feedback from the tool's users was positive, feedback highlighted wishes for expanding the tool. Therefore, a second project to develop an enhanced version of the tool, the "RAMP 2.0" was started with the objectives to: *i*) enhance the tools application range, to mainly also include work with repeated force exertion by the hand or fingers, *ii*) include key performance indicators (KPIs) as a support for managements and *iii*) in addition to the Excel-based version of the tool, also provide a web-based version of it.

A preliminary version of the RAMP 2.0 was launched in 2024 where the abbreviation changed from including 'manual handling' to include 'manual work' to reflect the tools enhanced application range. However, the web-based version did not function satisfactory, it needed more development work and was not released as planned. There was also a strong desire from users to include the novelties of RAMP 2.0 in RAMP courses.

The objective of this paper is to describe the new 'RAMP Package 2.0', consisting of *i*) the RAMP 2.0 tool, and present its new web-based version, *ii*) expanded courses on the tool and *iii*) the website, and in addition, present results from two evaluations of the tool.

Methods

The development of RAMP tools 2.0 version was carried out with a participative (Wilson, 1991), iterative (Martin et al., 2008) approach, similar to the methodology used for developing the first launched version of the tool, including literature studies, iterative development in close collaboration with intended users and collaborating with different expert groups connected to the project (Rose et al., 2020). The web-based version was developed in collaboration with in-house expert database developers. During the second project, courses of the RAMP were further developed using a smaller course development team than the first courses. The website was refined using user feedback. RAMP was evaluated in a user survey and in a Masters' Thesis.

Results

The RAMP Package 2.0 includes the RAMP 2.0 tool's four modules: 1) RAMP I for screening of MSD risks; 2) RAMP II for a more in depth analysis of the MSD risks, both I and II include the new Hand Model; 3) a new version of the Action Module to support development of risk reduction measures and action plans, including follow-ups and; 4) an improved Risk Management Support Module, consisting of three parts: a 'Process Description' on risk management processes and on the risk management parts of the RAMP tool, 'Aggregated Results' where results from different risk assessments can be aggregated and presented at different level of detail, and 'KPIs', which can be used to follow effects of using the tool. The new web-based version of RAMP 2.0 complements the Excel versions of the tool and enables more efficient analyses and aggregation and extraction of the risk assessments etc.

The RAMP Package 2.0 also includes two types of differentiated courses: In addition to the three existing fully online MOOCs open for all, three university courses with prerequisites to be accepted to the courses with university credits (ECTS) were developed, to address different users' needs. New material related to the tools '2.0' version was added to the website. Analysis of the website showed that the tool had been downloaded from 114 countries during its first seven years and done so by around 4000 unique users.

The results of a survey disseminated to all who had downloaded the tool during the first 26 months after its launch in 2017 showed that among the respondents around 1/2 used the tool, around 1/3 planned to do so and around 1/6 had at the moment not a plan to use it (Eriksson et al., forthcoming). They also showed that those who used the tool stated that using the tool had led to better work environment and lower rates of risks. In another study (Burghol, 2023) the reliability of RAMP 2.0's Hand model was evaluated; the new Hand Model had fair inter-rater and moderate intra-rater reliability. The results were similar to those of HAL TLV (Armstrong, 2006, Latko et al., 1997), which also was included in the study.

Discussion

There are many challenges when trying to implement MSD risk management tools as the one described here. A mayor one for companies is to be able to do this successfully – when they do not have a well-functioning *process* for the MSD risk management in place and yet want to use such tools for creating decent, sustainable jobs. To manage this difficulty is one of the UN's Sustainable Development goals (UN, 2015).

Key takeaways from the conference presentation

The 'RAMP Package 2.0' is renewed and enlarged and is now ready to be learned and used.

Acknowledgements

Financial support from AFA Insurance is gratefully acknowledged, as are contributions from companies, intended users, and research colleagues.

References

Armstrong, T.J. (2006). The American Conference of Governmental Industrial Hygienists threshold limit value for hand activity level. In: Marras, W.S., Karwowski, W. (Eds.), Fundamentals and Assessment Tools for Occupational Ergonomics. CRC Press, Boca Raton (FL), p. 41, 1-41.14.

Burghol, D. (2023). Evaluation of the RAMP Tool's New Hand Model-Reliability, Usability and Face Validity. Master Thesis, Royal Institute of Technology, Stockhpolm, Sweden.

Eriksson, A, Kluy, L, Rose, L (Forthcoming). Perceived usability and effects of using an MSD risk management tool.

European Agency for Safety and Health at Work (EU-OSHA) (2019). Work-related Musculoskeletal Disorders: Prevalence, Costs and Demographics in the EU. European Risk Observatory Report, ISSN, pp. 1831-9343.

Latko, W.A., Armstrong, T.J., Foulke, J.A., Herrin, G.D., Rabourn, R.A., Ulin, S.S. (1997). Development and evaluation of an observational method for assessing repetition in hand tasks. Am. Ind. Hyg. Assoc. J. 58 (4), 278-285.

Martin, J.L., Norris, B.J., Murphy, E., Crowe, J.A. (2008). Medical device development: the challenge for ergonomics. Appl. Ergon. 39 (3), 271-283.

Rose, L. M., Eklund, J., Nord Nilsson, L., Barman, L., Lind C. M. (2020). The RAMP package for MSD risk management – A tool and support for actions. *Applied Ergonomics*, 8, July 2020, 103101. https://doi.org/10.1016/j.apergo.2020.103101.

Rose, L. M., Neumann, W. P., Hägg, G. M. Kenttä, G. (2014). Fatigue and recovery during and after static loading. *Ergonomics*, 57(11), 1696–1710.

Sobhani, A., Wahab, M. I. M. & Neumann, W. P. (2015). Investigating Work-Related Ill Health Effects in Optimizing the Performance of Manufacturing Systems. *European Journal of Operational Research 241,* 708-718.

United Nations (2018). Sustainable development goals. Cited 2024, November 28. Available from: https://www.un.org/sustainabledevelopment/sustainable-development-goals/.

United Nations (2015). Transforming our world: the 2030 Agenda for Sustainable Development. Resolution adopted by the General Assembly on 25 September 2015. A/RES/70/1.

Wilson, J.R. (1991). Participation-a framework and a foundation for ergonomics? Journal of Occupational and Organizational *Psychology,* 64 (1), 67-80.

Using wearable sensor technology to improve learning transfer in manual handling training

Victoria L Forrester & Victoria Filingeri

University of Derby

SUMMARY

This research examined the use of wearable technology as an aid in manual handling training. The purpose of which was to establish if using wearable sensor technology can increase the transferability of training to the working environment and the factors that influence the likelihood of learning transfer taking place. Thematic analysis was used to examine the factors that may influence the transfer of learning. The findings suggest that wearable sensor technology has a place in aiding the transfer of learning from manual handling training.

KEYWORDS

Manual Handling Training, Wearable Sensor Technology, Learning Transfer

Introduction

The purpose of this research was to examine if using wearable sensor technology in manual handling training aids the transfer of learning to real working environments. The study aimed to investigate the factors that influence transfer of learning. Manual handling has been identified as a key contributor to work related musculoskeletal disorders (MSDs) across a range of industry populations including healthcare, manufacturing, fisheries, logistics and agriculture (Bork et al., 1996; Deros et al., 2010; Basahel, 2015; Asuquo et al., 2021; Remmen et al., 2021; Das, 2023). Despite the multicausal nature of work-related MSDs manual handling training is the primary method of mitigating and reducing the risk of an employee developing MSDs from their work activities. Whilst the use of manual handling training is extensive many researchers suggest that it is ineffective in reducing work related musculoskeletal injuries (Haslam et, al. 2007; Martimo et al. 2007; Clemes et al. 2010). Denis et al. (2020) reinforces this opinion by describing most workplace manual handling training as being delivered in a parachute fashion in a class room setting whereby the training is standardised regardless of the working environment. The training invariably involves lifting a standard box which does not relate to real work settings, and which consequently may result in limited transfer of learning.

Lind et al., (2023) note that there is a growing use of wearable sensor technology within the field of ergonomics. Wearable sensor technology uses multiple sensors connected to an external receiver such as a tablet device to objectively record biomechanical movement and complete a risk analysis on the data obtained. It enables participants to objectively observe their manual handling technique and in doing so highlight high risk biomechanical movement. Spook et al., (2019) suggest that monitoring work exposure with wearable sensor technology may have the potential to promote worker health. While Lind (2023) perceives that wearable sensor technology has a potential role in preventing work related MSDs and thus maybe beneficial in improving the effectiveness of manual handling training. Due the gap in the existing research this study aimed to examine the use of wearable sensor technology to improve the transfer of learning in manual handling training.

Methods

Employees perceptions provided the framework for examining the use of wearable sensor technology in manual handling training. The participants were 10 male warehouse operatives (22-59 years), who are required to undertake manual handling activities as part of their work role. The initial part of the study was the delivery of a Power Point presentation around manual handling techniques and the risk associated with manual handling. Following this education session non-invasive wearable sensors were used to obtain biomechanical data to analyse work tasks and in turn a risk rating was determined. Joint Action Solutions (JAS) was the brand of wearable sensor technology used within this study. JAS is an automated injury risk assessment system. The system uses non-invasive wearable sensors that obtain musculoskeletal data to analyse tasks and generate scored manual task risk assessments delivered to an iPad. JAS uses a machine processing system to process the sensor data. Sensors are placed on both arms over clothing via an elasticated strap / band. A waist sensor was attached to a belt that the participants were asked to wear. A sensor was also placed via a clip to the back of the participant's collar. The final sensor was placed at the back of the head which is held in place through a pocket in an elasticated head band. There is no direct skin contact with the sensors. The parameters measured were as follows; neck flexion/extension, rotation and side flexion; back flexion/extension, rotation and side flexion; arm left arm elevation and right arm elevation.

Individually each participant completed a work manual handling task which was videoed for between 5 and 8 minutes on an iPad. The use of the video allowed the data to be collected and viewed in context. Following data collection participants were given feedback around their manual handling lifting technique with good practice reinforced. The feedback was verbal with the video recording being used to highlight the movements that had the potential to create musculoskeletal problems. Participants were asked to implement any new learning into their daily work based manual handling tasks. A follow up semi structured questionnaire was conducted four weeks after the training, in order to ascertain the participants views on the benefits and effectiveness of the training provided. The interviews were transcribed from which key themes were identified.

Analysis and Results

Thematic analysis (Braun & Clarke 2022) was the analytical strategy used to assess the data. Through thematic analysis overarching themes were identified with seven sub themes. The themes and subthemes are shown in figure 1.

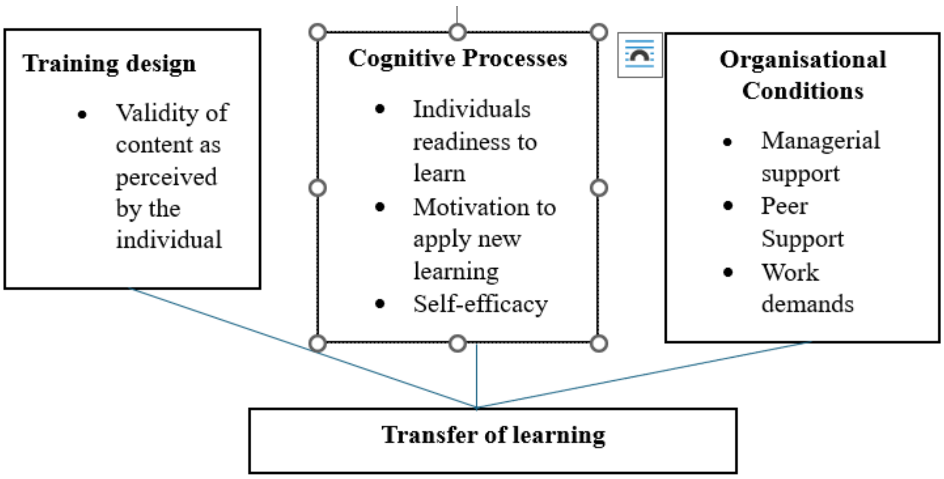


Figure 1: Key theme and sub themes drawn from the interview transcripts.

The research analysis showed that participants previous experiences was very generic in nature whereas the use of wearable sensor technology was more content relevant to their work role and thus easier to apply in daily manual handling tasks which they perceived added validity to the training. Content validity is seen as an important factor in the transfer of learning. Content validity is viewed as the relatability of the manual handling training to the real working environment.

Cognitive processes were also a determining factor in the participants perception of the learning obtained and the transfer of the learning to their work tasks. Cognitive processes are considered as the mental action or process of acquiring knowledge and understanding thoughts and experiences. It encompasses processes such as perception, thought, attention, intelligence and imagination in the formation of knowledge, the ability to make decisions, memory creation, and the ability to evaluate and create reasoning (Smid et al., 2020). Cognitive processes inform decision making and in the case of this research this relates to transferring the learning gained from the use of wearable sensor technology in manual handling training. Cognitive process subthemes included a participant's readiness to learn, their motivation to apply new learning and self-efficacy.

The final theme that emerged was organisational conditions**.** Managerial support, peer support and work demands were viewed and factors in the transfer of learning.

Conclusion

This research examined the use of wearable sensor technology and its value in transferring manual handling training to a real-world occupational setting. Wearable sensor technology has been used to assess the potential for work related MSD. The advancement of technology and it's use within the field of ergonomics affords the opportunity to explore if technology such as the wearable sensor technology used in this research can be employed as a training aid which can ultimately improve the transfer of learning. The use of wearable sensor technology in manual handling training improves perceived content validity due to its relevance to the participant. The challenge moving forward is to embrace technology and explore where it can be beneficial in reducing work-related MSDs.

References

Asuquo, E. G., Tighe, S. M., & Bradshaw, C. (2021). Interventions to reduce work-related musculoskeletal disorders among healthcare staff in nursing homes; An integrative literature review. *International Journal of Nursing Studies Advances*, *3*, 100033

Basahel, A. M. (2015). Investigation of work-related musculoskeletal disorders (MSDs) in warehouse workers in Saudi Arabia. *Procedia Manufacturing*, *3*, 4643-4649.

Bork, B. E., Cook, T. M., Rosecrance, J. C., Engelhardt, K. A., Thomason, M. E. J., Wauford, I. J., & Worley, R. K. (1996). Work-related musculoskeletal disorders among physical therapists. *Physical therapy*, *76*(8), 827-835.

Braun, V., & Clarke, V. (2022). Conceptual and design thinking for thematic analysis. *Qualitative Psychology*, *9*(1), 3.

Clemes, S. A., Haslam, C. O., & Haslam, R. A. (2010). What constitutes effective manual handling training? A systematic review. *Occupational medicine*, *60*(2), 101-107

Das, B. (2023). Work-related musculoskeletal disorders in agriculture: Ergonomics risk assessment and its prevention among Indian farmers. *Work*, (Preprint), 1-17.

Denis, D., Gonella, M., Comeau, M., & Lauzier, M. (2020). Questioning the value of manual material handling training: A scoping and critical literature review. *Applied Ergonomics*, *89*, 103186.

Deros, B. M., Daruis, D. D., Ismail, A. R., Sawal, N. A., & Ghani, J. A. (2010). Work-related musculoskeletal disorders among workers' performing manual material handling work in an automotive manufacturing company. *American Journal of applied sciences*, *7*(8), 1087

Haslam, C., Clemes, S., McDermott, H., Shaw, K., Williams, C., & Haslam, R. (2007). Manual Handling Training Investigation of Current Practices and Development of Guidelines Prepared by the Work and *Health Research Centre Investigation of Current Practices and Development of Guidelines*. [online][Accessed May 2024].

Lind, C. M., De Clercq, B., Forsman, M., Grootaers, A., Verbrugghe, M., Van Dyck, L., & Yang, L. (2023). Effectiveness and usability of real-time vibrotactile feedback training to reduce postural exposure in real manual sorting work. *Ergonomics*, *66*(2), 198-216.

Martimo, K. P., Verbeek, J. H., Karppinen, J., Furlan, A. D., Kuijer, P. P. F., Viikari-Juntura, E., ... & Jauhiainen, M. (2007). Manual material handling advice and assistive devices for preventing and treating back pain in workers. *Cochrane Database of Systematic Reviews*, (3).

Remmen, L. N., Heiberg, R. F., Christiansen, D. H., Herttua, K., & Berg-Beckhoff, G. (2021). Work-related musculoskeletal disorders among occupational fishermen: A systematic literature review. *Occupational and environmental medicine*, *78*(7), 522-529.

Smid, C. R., Karbach, J., & Steinbeis, N. (2020). Toward a science of effective cognitive training. *Current Directions in Psychological Science*, *29*(6), 531-537.

Spook, S. M., Koolhaas, W., Bültmann, U., & Brouwer, S. (2019). Implementing sensor technology applications for workplace health promotion: A needs assessment among workers with physically demanding work. *BMC Public Health*, *19*, 1-9.

What can the Post Office 'Horizon' scandal teach about Artificial Intelligence deployment?

Chris Baber[1], Brandon King[2], Paul Salmon[2] and Yihao Jiang[1]

[1]University of Birmingham, UK, [2]University of Sunshine Coast, Australia

SUMMARY

In this paper we use the Post Office Horizon scandal. Broadly, accounting errors arising from a computer system were falsely attributed to Subpostmasters who were subsequently accused of theft, fraud, and false accounting. Over 700 prosecutions have been successfully appealed and convictions quashed. Coupled with severe technical shortcomings, there was confirmation bias in the decisions of Post Office Limited to prosecute Subpostmasters and an assumption in UK Law that computers do not make mistakes (unless proven otherwise). From a review of evidence and newspaper reports of specific cases in the Post Office Horizon scandal, we construct and analyse Accimaps. We argue that a common problem across these cases is how different sorts of 'black box' of the Horizon system meant that it lacked transparency for all stakeholders.

KEYWORDS

Post Office Horizon; Accimaps; Socio-Technical Systems; Transparency

Introduction

In the UK, a public enquiry into problems arising from the 'Horizon' computer system has been widely reported. These problems were brought to the public's attention via a BBC television programme 'Mr. Bates vs the Post Office' in 2023. In the UK, all Post Office branches are franchised by Sub post-masters (SPM) from Post Office Limited (POL). Most SPMs also run other businesses from their Post Office, e.g., a general store or newsagents etc. This means that Post Office branches might have two sets of tills and two sets of accounts, only one of which is returnable to POL. Further, the franchise agreement states that SPMs are responsible for any shortfalls in returns to POL. This means that if the return is incorrect, the SPM will need to make up the missing money (there does not seem to be a requirement to handle over-payment).

In the Horizon scandal, over 730 SPMs were prosecuted by POL between 2000 and 2014. Many of these were accused of theft, which could lead to a prison sentence, and were offered a plea deal of 'false accounting', a lesser charge, and repay the 'missing' money. Most accepted this deal. It is worth noting that POL does not have any special powers under UK law, but (like any other citizen) can bring its own private prosecutions. In contrast to prosecutions brought by the Police, private prosecutions do not require scrutiny by the Crown Prosecution Service[1]. Where POL differed from other citizens it the resources that it could apply to this process and the fact that it employs its own investigators. External auditing of accounts has not received much attention to date. In many instances, SPMs agreed to settle out of Court and pay the outstanding balance (and in the cases that have been reported, SPMs had been paying smaller shortfalls from their own money prior to being brought to Court).

[1] https://www.gov.uk/government/consultations/oversight-and-regulation-of-private-prosecutors-in-the-criminal-justice-system/oversight-and-regulation-of-private-prosecutors-in-the-criminal-justice-system-consultation

When computers were first considered in relation to crime (in the late 1970s) computer evidence was treated as 'hearsay' and, hence, inadmissible. In section 69 of the 1984 Police and Criminal Evidence Act (PACE), there was a requirement for the prosecution to prove that the computer was working correctly before evidence could be admitted. Checking all computers related to a crime was burdensome, so section 69 of PACE was repealed (under section 60 of the Youth and Criminal Evidence Act in 2000). Thus, the presumption in English law is that a computer is 'working properly'. It is the burden of the person challenging evidence to prove that the computer was not working properly. In most of the *Horizon* cases, Courts assumed that the computer was working properly, and the Defence was not given access to documents from POL or Fujitsu that could allow this assumption to be challenged, and (except in a couple of cases) the Court would not accept a claim that the computer had malfunctioned.

In his instructions to the jury in the case of Seema Misra in 2010, the judge asked, "Do you accept the prosecution case that there is ample evidence before you to establish that Horizon is a tried and tested system in use at thousands of post offices for several years, fundamentally robust and reliable?" Misra pleaded guilty to six counts of false accounting and the jury found her guilty of stealing £74,000 (Brooks and Wallis, n.d.). This verdict was given despite the judge's observation that, "There is no direct evidence of her taking any money...There is no evidence of her accumulating cash anywhere else or spending large sums of money or paying off debts, no evidence about her bank accounts at all. Nothing incriminating was found when her home was searched."

From this brief introduction, it should be apparent that we are considering a Sociotechnical System in which technology (*Horizon*) is embedded in an organizational culture (in POL) that was, at best, suspicious of the accounting of SPMs. The Horizon scandal was compounded by a legal system that presumes computer technology is reliable and by denials by POL that there were problems with *Horizon*. As early as 2003, in the trial of Julie Wolstenholme, the judge accepted expert witness testimony that *Horizon* was faulty. In response, POL stated "It is denied that the said computer system was unfit for its purpose and it is averred that the same worked adequately."[2]

Perrow (1990) suggested that any reasonably complex system is always in a near accident state. Such states are kept in check because safeguards are in place to limit potential risks, or because people can monitor and intervene to control the system, or because regulations exist to penalize the outcomes of critical states. Sometimes all three fail and the Horizon scandal is an example of this. In terms of system failures, one might expect the 'system' under consideration to be sufficiently transparent for stakeholders to understand and manage (Woods, 1996). An initial step in an Ergonomics analysis of the Horizon scandal would be to ask (a) what defined the boundaries of the system under consideration, and (b) was the system transparent to its operators and stakeholders.

In this paper, we consider the boundaries of the system through the lens of Accimaps, and transparency in terms of 'black boxes'. In computing, a 'black box' is a process that is opaque to its users, typically because of its complexity, and the user will be expected to trust the output of this process. The black box problem is increasingly pertinent with the application of Artificial Intelligence, AI (von Eschenbach, 2021) and we believe that the problems relating to a non-AI computers in a Sociotechnical System can illustrate the challenges that AI might pose.

A brief overview of the UK Post Office Horizon Computer System

International Computers Limited (ICL) was awarded a contract from UK Government in 1996 for a computer system to support a Benefits Payment Card for payment of unemployment and other

[2] https://www.computerweekly.com/news/366546032/Post-Office-tried-to-convince-independent-IT-witness-that-he-was-wrong-about-Horizon [accessed 12th February 2025]

benefits by Post Office branches. This project, *Pathway*, was discontinued due to "greater than expected complexity", with the House of Commons public accounts committee calling it "one of the biggest IT failures in the public sector". However, in 1999 ICL signed a contract to automate Post Office operations by the Pathway team and this resulted in *Horizon*. Fujitsu had bought 80% of shares in ICL by 1990 and by 1998 was the sole shareholder. *Horizon* proved to be a lucrative contract for Fujitsu. It is estimated that Fujitsu has received around £2.5 billion[3] for Horizon operations. In addition to a fixed payment for delivering the system (and upgrading it in 2021, 2023 and 2024), Fujitsu receives payment for each of the 2 million+ transactions processed each day.

Horizon integrated systems and services from a variety of providers, e.g., Oracle, Escher, ICL, ATOS, Computacenter, using a variety of programming languages, e.g., VisualBasic, C, and C++, and databases built in Oracle. Thus, *Horizon* was an amalgamation of sub-systems, combined using a variety of languages and built on the already discredited *Pathway* system. One might anticipate that, without appropriate testing, integration would be problematic. To make matters worse, the development team for the project was small and not familiar with the nuances of the various systems. A member of the development team, David McDonnell told the 2024 public inquiry "of eight [people] in the development team, two were very good, another two were mediocre but we could work with them, and then there were probably three or four who just weren't up to it and weren't capable of producing professional code".[4]

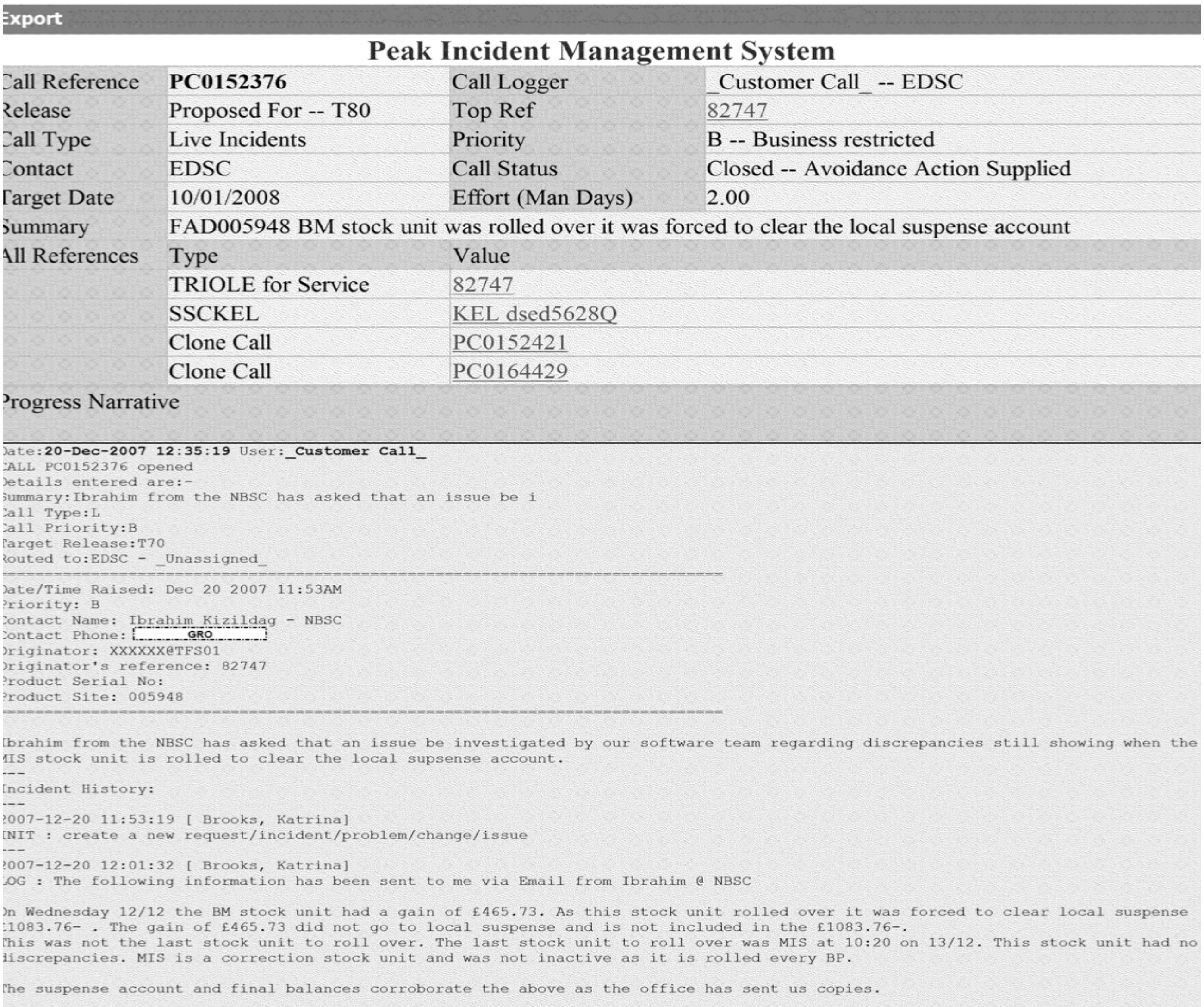
For any database transaction, one can apply four criteria: Atomicity (statements are treated as single units to be executed); Consistency (changes to the database must be predictable); Isolation (transactions from different users of the database need to be kept separate); Durability (changes to data will be saved, even if the system fails). The examples considered below demonstrate failures of atomicity (because the transactions were duplicated); consistency (the examples suggest an inability to synchronize the various parts of the system, which led to mismatches); isolation (the examples suggest that operations had the potential to interfere with each other); durability (some entries (and errors) could not be un-done). A detailed discussion of these issues can be found online.[5] As noted previously, *Horizon* was an amalgamation of systems that were integrated together and combined with the poorly built *Pathway* system. A key issue with *Horizon* was how it managed data. While there was an Oracle database, much of the data management used an XML structure that wrote messages to a message store and there was no agreed catalogue or dictionary of the message, so no attempt to ensure their consistency (Wallis, 2021).

Horizon included an electronic point of sales (EPOS) terminal and accounting transaction management system for Data Reconciliation between the Post Office branch and POL baock-office computers. Connection between branches and back-office was via a telephone line (until the introduction of Horizon Online). This would inevitably cause delays in transmission of data, particularly in branches that were in remote locations. As an accounting system, *Horizon* ran according to double entry book-keeping. When the SPM was recording a transaction, this involved two double entry baskets: Transfer Out from one Stock Unit and Transfer In to another Stock Unit. When a Transfer was made, then an indicator of outstanding transfers should be updated. But, this indicator was not displayed on all EPOS terminals so the SPM would not know if the Transfer had been made, so might attempt to repeat the Transfer. Balancing of the accounts assumed that there were no outstanding transfers (that is, it requires Transfer Out to equal Transfer In). However, if (as happened in the Callendar Square 'bug', mentioned below) there were 2 Transfers In for a single

[3] https://www.computerweekly.com/news/366586814/Post-Office-Horizon-replacement-project-labelled-unachievable-as-taxpayer-bill-reaches-1bn
[4] https://www.theguardian.com/uk-news/2024/jan/09/how-the-post-offices-horizon-system-failed-a-technical-breakdown [accessed 31st January 2025]
[5] https://evidencecritical.systems/2021/07/15/what-went-wrong-with-horizon.html [accessed 30th January 2025]

Transfer Out, then this affects the cash held in the branch and is defined as a loss. According to Wallis (2021) “The cash account wasn’t an account in the traditional sense of the word. It was a program which crawled through every transaction on each Horizon terminal in each branch at the end of the day’s trading. It then came up with a figure which should correspond exactly with the amount of physical cash on the premises. That figure was then automatically uploaded to the Post Office’s central servers overnight. It is a relatively simple task to describe, but not necessarily to execute. Given its central importance to the financial integrity of the Horizon system, it had to be bullet-proof. It wasn’t. The code was not good enough.” (p.12)

From the early deployment, the development team maintained a log of ‘bugs’ (although a full list of these has not been disclosed). Fujitsu engineers maintained a Known Error Log (KEL) and the PEAK incident management system. An example of a PEAK record is shown in figure 1.



Export

Peak Incident Management System

Call Reference	**PC0152376**	Call Logger	_Customer Call_ -- EDSC
Release	Proposed For -- T80	Top Ref	82747
Call Type	Live Incidents	Priority	B -- Business restricted
Contact	EDSC	Call Status	Closed -- Avoidance Action Supplied
Target Date	10/01/2008	Effort (Man Days)	2.00
Summary	FAD005948 BM stock unit was rolled over it was forced to clear the local suspense account		
All References	Type	Value	
	TRIOLE for Service	82747	
	SSCKEL	KEL dsed5628Q	
	Clone Call	PC0152421	
	Clone Call	PC0164429	

Progress Narrative

```
Date:20-Dec-2007 12:35:19 User:_Customer Call_
CALL PC0152376 opened
Details entered are:-
Summary:Ibrahim from the NBSC has asked that an issue be i
Call Type:L
Call Priority:B
Target Release:T70
Routed to:EDSC - _Unassigned_
================================================================================
Date/Time Raised: Dec 20 2007 11:53AM
Priority: B
Contact Name: Ibrahim Kizildag - NBSC
Contact Phone: GRO
Originator: XXXXXX@TFS01
Originator's reference: 82747
Product Serial No:
Product Site: 005948
================================================================================

Ibrahim from the NBSC has asked that an issue be investigated by our software team regarding discrepancies still showing when the
MIS stock unit is rolled to clear the local supsense account.
---
Incident History:
---
2007-12-20 11:53:19 [ Brooks, Katrina]
INIT : create a new request/incident/problem/change/issue
---
2007-12-20 12:01:32 [ Brooks, Katrina]
LOG : The following information has been sent to me via Email from Ibrahim @ NBSC

On Wednesday 12/12 the BM stock unit had a gain of £465.73. As this stock unit rolled over it was forced to clear local suspense
£1083.76- . The gain of £465.73 did not go to local suspense and is not included in the £1083.76-.
This was not the last stock unit to roll over. The last stock unit to roll over was MIS at 10:20 on 13/12. This stock unit had no
discrepancies. MIS is a correction stock unit and was not inactive as it is rolled every BP.

The suspense account and final balances corroborate the above as the office has sent us copies.
```

Figure 1: extract from PEAK PCO152376

Detica[6] identified four areas of risk (in terms of operational practice) in POL operations: “…non-conformance to Post Office policy and processes by branches, with an institutionalised acceptance that errors, workarounds and non-conformance exists; Complexity and fragmentation of information systems which hamper efforts both to gain an insight into branch behaviour and root

[6] https://www.jfsa.org.uk/uploads/5/4/3/1/54312921/document_25_-_detica_netreveal_fraud_analysis_011013_1.pdf [accessed 6th February 2025]

causes; Ineffective process, policy and working practice in the central operational teams to gather information, prioritise and act in a co-ordinated manner; Technology available to central operational teams are not fit for purpose; analysis of large data sets is performed on an ad-hoc basis of data subsets copied into Excel and tasking of teams is initiated and managed through email."

The first point is of interest to this paper because it highlights the complexities in defining a single root cause of the 'technical' issues relating to *Horizon.* The Detica report provides the example of selling non-Post Office products through a Post Office till. As noted previously, many Post Office branches are run in conjunction with other shops. While the policy might be that only Post Office products (stamps, postage etc.) will be put through the *Horizon* EPOS terminal, it was common for SPMs to also record transactions of items from their other shop (e.g., greetings cards) through the EPOS and to record these as 'Postage other'. This transaction would then be reversed at the end of the day (or the accounting period). As the Detica report notes, "Instances of non-conformance also generate operational noise which hides deliberate attempts to defraud it. This has resulted in a large number of false positives when looking for fraud, and inhibits Post Office's ability to detect fraud early, resulting in larger losses." Thus, dealing with 'deliberate fraud' was a challenge for POL. Given the prevalence of multiple businesses being from Post Office branches, POL might have assumed that SPMs were using money to cross-subsidise these other businesses in Post Office branches. This could have led to the belief that the scale of this 'theft' was revealed by *Horizon.* Consequently, there may have been a degree of confirmation bias behind POL prosecutions of SPMs.

Technical Failings of Horizon

Descriptions of some technical failures are provided in the Technical Appendix to 'Alan Bates and others and the Post Office Limited'. POL argued that there were no technical problems with *Horizon.* At first, 'bugs' were named after the Post Office branch in which they were reported. This naming convention suggests that any problem is due to the branch rather than the computer system. This might also reflect Fujitsu's assumption that the 'bugs' were due to human error rather than technical failings. As Graham Ward, Fujitsu security case manager, commented in an email to staff at Fujitsu about a report on Noel Thomas (who was jailed for £48k alleged false accounting) "Given the allegations made by the postmasters, I'm sure you'll agree that it's very much in ourselves and Fujitsu's interest to challenge the allegations and provide evidence that the system is not to blame for the losses provided."[7]

Callendar Square and Dalmellington bugs

The Dalmellington bug occurred when a user entered a transaction of £8000 on the EPOS terminal. The screen froze, so the user hit the 'enter' key again – duplicating the entry. With second and third presses of the enter key, the transaction accumulated to £32,000. As noted above, the SPM was liable for the discrepancy of £24,000, which was paid from personal funds. The incident was raised in PEAK PCO126042 (15/09/05) which also noted an instance recorded in 2000. Related issues were noted in PEAKS PC0056922, PC0075 92, PCO083101, PCO193012, PCO103864. KEL JSimpkins338Q, May 2002 seems similar and describes the root cause as a Riposte error "Timeout occurred waiting for lock (0xC1090003)". Riposte was supplied by Escher and, in 2006, the problem was fixed by Escher, although there is a Peak (PC0193012 - referenced by JSimpkins338Q - which was raised in 2010 and again a reboot fixed the issue, and these PEAKs closed in 2010.

[7] https://www.computerweekly.com/news/366589716/Metropolitan-Police-set-to-investigate-one-of-its-own-staff-in-Post-Office-probe#:~:text=In%20the%20email%20to%20Fujitsu,his%20edit%20of%20the%20statement [accessed 2nd February 2025]

Callendar Square is reported in the following PEAKS: Peak PCO126042 raised 15/09/2005; Peak PCO126376 raised 21/09/2005; PC0056922; PC0075 92; PCO083101; PCO193012. It is reported in KEL JBallantyne5245K, KEL JSimpkins338Q (from 2002). The root cause was a time-out bug in Riposte that prevented EPOS terminal from writing messages.

Accimaps

We constructed Accimaps of individual cases using source materials from the public enquiry and court cases, supplementing these with newspaper and magazine articles, journal papers, and opinion pieces of websites. Originally developed by Rasmussen (1997), Accimaps visualize the interplay of factors that contribute to an incident at levels of Socio-Technical System, i.e., wider society, legal and regulatory, organizational, technical, individual. These are a popular approach to accident analysis (Salmon et al., 2023). We are interested in how technical failings of *Horizon* (which are not always well documented) contribute to the scandal.

Accimap of a known technical failure: Girobank reconciling

In PEAK PC0044232 (KEL_MWright531p) there was a £505.72 discrepancy. There seem to be five complicating factors: (i.) this involved a giro (which is like a cheque, in this instance one issued by the UK benefits agency) which is cashed by the SPM for a customer after the last collection of giros from the branch (every day Girobank would send a van around the post offices to collect any giros that had been cashed, take these back to Girobank to process, and then Girobank would pay the post office back the money that had been paid out for each giro); (ii.) because the SPM knew that the last collection had been missed, he cancelled the giro transaction after entering it in *Horizon*, intending it to be recorded the following day after the giro had been collected; (iii.) Girobank receive the giro but also notification from Horizon that the transaction had been cancelled, so their system issues an error notice; (iv.) the SPM should (according to company policy) reject the error notice so that the amount of cash can be recorded in Horizon correctly; (v.), the SPM did not know this until the end of the Trading Report period (which was every week - not every day).

Figure 2 illustrates how the Horizon system (in addition to technical problems) operated in a wider technical system in which the timing of events and interconnection of different processes created a complicated set of relations. In order to deal with these complications, heuristics were applied and these included counter-intuitive tasks, such as cancelling and then correcting the Girobank error notice, which needed to be performed at a different time to the original transaction (in this case, several days after the transaction).

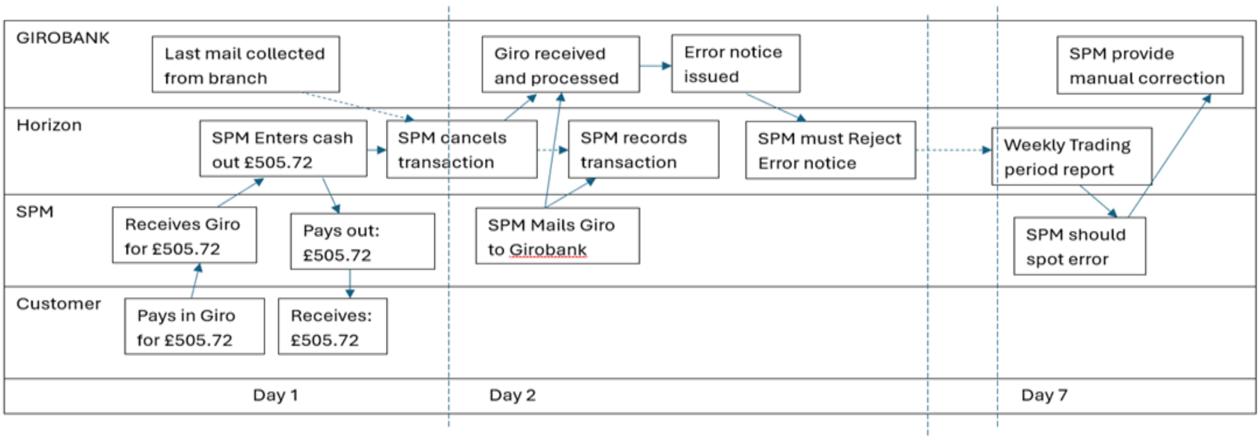


Figure 2: Accimap of Girobank discrepancy issue

Accimap of wider systemic failure: The case of Jo Hamilton

Jo Hamilton was featured in the BBC 'Mr. Bates vs the Post Office' and her case is sufficiently well reported to allow us to construct the Accimap (a section of this is shown in figure 3). She became SPM in 2003. *Horizon* showed shortfalls in the accounts. Assuming that these were the result of problems that arose from her mistakes in using the system, she paid the shortfalls from her own money. From 2003 to 2006 the errors accumulated and she re-mortgaged her house to continue making payments. In 2006, the discrepancy was over £10,000 and she was suspended. When Hamilton was suspended, she was not allowed back into the Post Office branch. This meant that she and her defence struggled could not access the branch computer. It is plausible to assume that, even if such access was granted, the records of transactions might not have been stored in a reliable manner. Consequently, the prosecution case relied on the assumption of a correctly working computer. In 2008 she was taken to Court and charged with the theft of £36,644.89. To avoid a prison sentence, she agreed to a plea bargain in which she pleaded guilty to 14 counts of false accounting. She had to remortgage her house again to pay the outstanding amount plus £1000 towards prosecution costs. She was also sentenced to a community order.

The initial problems in Jo Hamilton's case appear similar to those of the Dalmellington 'bug' (although there is a lack of technical detail both in Hamilton's accounts of the incidents and the Post Office's prosecution statements). Hamilton sought assistance (through hundreds of calls) from the help-desks offered by Post Office Limited. There were at least 3 levels over which the help desks operated:

Level 1: HSH is Fujitsu Horizon Helpdesk – call handlers reply to SPM questions using POL operating procedures and gives advice from scripts. So, these are unlikely to help in solving a problem. There was also a National Business Support Centre which was POL helpdesk which responds to security problems, complaints about Horizon, other operating problems.

Level 2: SMC is Fujitsu Horizon unit that monitor the 'event storms' from counter (which I believe are the Electronic Point of Sales (EPOS) terminals in branches). These can access Known Error Logs (KELs) to find previous fixes. There was also (in the earliest deployment) the ICL Pathway System Support Centre.

Level 3: Fujitsu System Support Centre dealt with 'new' problems and logged these in a system called PINICL (which was later replaced by a system called PEAK).

The likely flow would be for the SPM to call a level 1 helpdesk and receive a standard answer, such as turn the system off and on again which appeared to solve some of the issues associated with bugs such as Dalmellington. This help was provided from a script and would not directly address specific problems. If the problem was not solved, then this might escalate to level 2 where the solution to the problem could be retrieved from a KEL. If this did not work then the problem might escalate to level 3 where, if there was not a KEL, would be recorded as a PEAK.

Conclusions

We propose lessons that can be learned from *Horizon* are applicable to the deployment of Artificial Intelligence (despite *Horizon* using poorly designed technology). The lessons relate to Sociotechnical systems and can be understood in terms 'black boxes' at different levels.

- For sub-postmasters *Horizon* was opaque and produced confusing and inconsistent output. This was compounded by poor user interface design, limited training, and unpredictable software. This resulted in a lack of transparency that made it impossible for the SPM to determine the source of a problem in the computer. In a few instances, the SPM kept a parallel record of transactions which could be used to challenge the *Horizon* output.

- In English Law, the assumption is that any computer is working correctly at the time of any incident, unless there is evidence to the contrary. This presumption means that the burden is on the defence to prove failure of a computer. Lack of access to computer logs for defence lawyers meant that *Horizon* was a black box for them. The manually kept records (noted above) were seldom sufficient to challenge the computer records in Court.

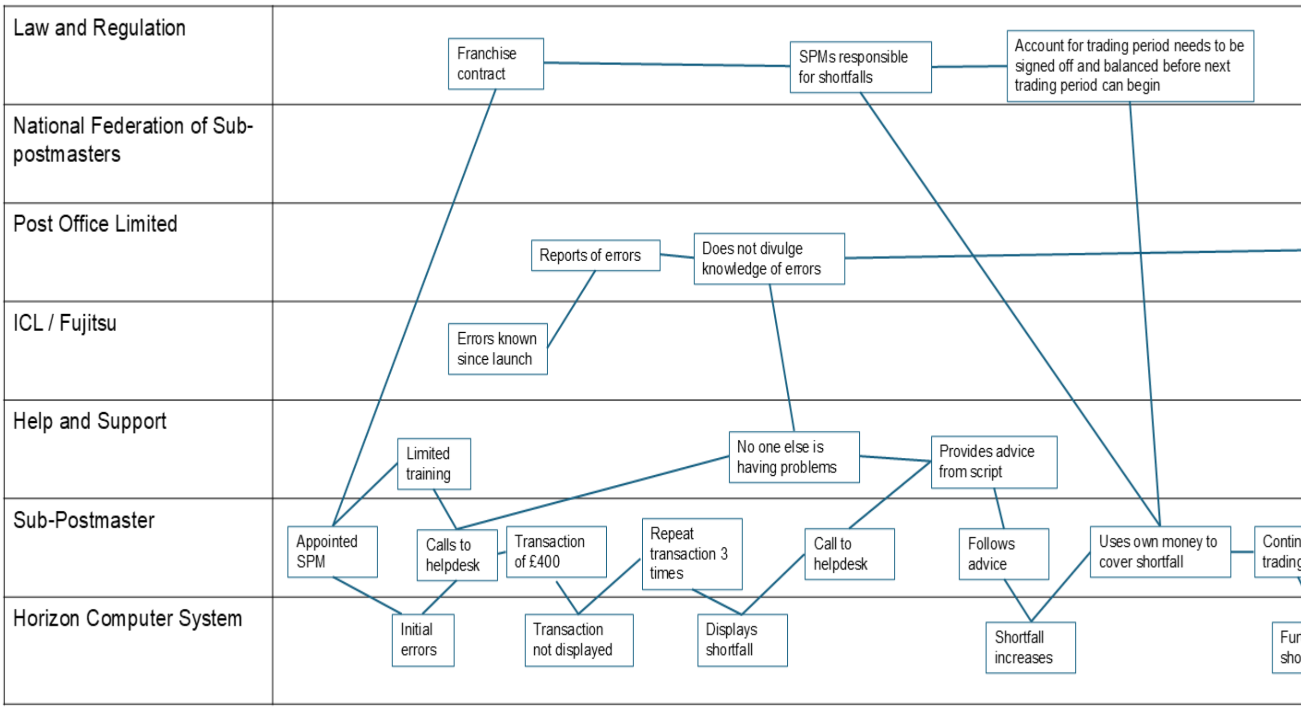


Figure 3: Extract of Accimap summarising Jo Hamilton’s incidents

- For system developers a combination of limited competence, integration of incomplete systems, and limited oversight of failures, meant that *Horizon* was a black box for the team who developed and deployed it. Fujitsu ran a team that attempted to correct errors through remote access to Post Office branch computers although, as late at 2015, POL told a House of Commons inquiry that “There is no functionality in Horizon for either a branch, Post Office or Fujitsu to edit, manipulate or remove transaction data once it has been recorded in a branch’s accounts.”
- For Post Office Limited there seemed to be a lack of knowledge or interest in the operation of the computer system and how it might not be performing correctly. This suggests a black box perspective of technology taken by management and their advisors. More pernicious is the implication that viewing the *Horizon* system as a black allowed the confirmation bias of assumed guilt of SPMs.

We propose that multi-level, multi-stakeholder black boxes are a critical issue in sociotechnical systems, and that the transparency issues they create play a causal role in adverse events. This work is critical as it points to the need for improved transparency in AI technologies, both in terms of the AI itself, the human-machine interface, and associated procedures, operating rules, and regulations.

References

Brooks, R. and Wallis, N., Justice lost in the post: how the post office wrecked the lives of its own workers, Private Eye, https://www.private-eye.co.uk/special-reports/justice-lost-in-the-post [accessed 30th January 2025]

Woods D. D. (1996). Decomposing automation: Apparent simplicity, real complexity. In Parasuraman R., Mouloua M. (Eds.), *Automation and human performance: Theory and applications* (pp. 3–17). Erlbaum.

von Eschenbach, W.J. Transparency and the Black Box Problem: Why We Do Not Trust AI. *Philos. Technol.* 34, 1607–1622 (2021)

Perrow, C., 1984, *Normal Accidents: living with high-risk technologies,* New York: Basic Books

Rasmussen, J., 1997, Risk management in a dynamic society: a modelling problem, Safety Science, 27, 183-213

Salmon, P.M., Hulme, A., Walker, G.H., Waterson, P., Berber, E. and Stanton, N.A., 2020. The big picture on accident causation: A review, synthesis and meta-analysis of AcciMap studies. *Safety science*, *126*, p.104650.

Wallis, N. (2021) *The Great Post Office Scandal*, Bath Publishing

Working at Height Assessment: Learning from Experience on Military Vehicles

Elaine McDonald

QinetiQ, UK

SUMMARY

Working at height introduces additional risks to individuals in the workplace. A major risk is that of falling. Falling from height is one of the leading causes of workplace fatalities. Falling from height results in high kinetic injuries; despite advances in medical care, these injuries have high morbidity. The Health and Safety Executive (HSE) currently provides guidance to practitioners on what needs to be done to protect individuals. The guidance was found to be not as extensive as initially expected and introduced uncertainty on the approach to take. It should be considered whether the working at height guidance could be expanded and how a more comprehensive guide could be created.

KEYWORDS

Working at Height, Guidance, Risk

Introduction

Falls from height are one of the leading causes of fatalities and major injuries in the workplace (HSE, 2014). In 2023/24, it was reported that 50 fatal injuries in Great Britain were due to falling from height (HSE, 2024). An increased morbidity is associated with falls from height due to the high kinetic nature of these injuries (Alizo et al 2017). As such, the risks and consequences of working at height should not be underestimated.

Working at height is a reality for many Armed Forces personnel when undertaking their normal roles and day-to-day tasks. Regardless of whether they are completing maintenance tasks, or out on operations, there is often a need for personnel to work on the roofs of land platforms. Example platforms heights are (Army, 2024a; Army 2024b; Defence Equipment & Support, n.d.):

- 2.78 metres for Warrior;
- 3.4 metres for Stormer; and
- 2.5 metres for Challenger 2.

The Armed Forces seek to abide to the HSE regulations, but the training for and deployment into combat environments during military operations pose unique challenges, where the falling from height risk increases. This paper looks at the HSE Guidance and the experience of applying that guidance to military vehicles.

HSE Guidance

Guidance for working at height is provided by the HSE, which takes the form of a brief guide. Within the guide a step-by-step diagram is provided flowing through the following decision steps:

- Can you AVOID working at height in the first place? If NO, got to PREVENT

- Can you PREVENT a fall from occurring? If NO, go to MINIMISE
- Can you MINIMISE the distance and /or consequences of a fall?

Applying HSE Guidance

What we did

A recent project required a working at height assessment on a military vehicle to understand whether the safety measures and provided fall arrest equipment were sufficient to protect those working at height on the vehicle.

At the outset of the task the locations at which personnel may be working at height and the tasks performed at those locations were identified. The HSE guidance was then reviewed and the following prompts were selected to guide in-person questioning during the vehicle working at height assessment:

1. How is the task completed?
2. What protections are in place for a fall from height?
3. What would be the impact of using fall arrest equipment on task performance?
4. Can you AVOID working at height? If NO go to PREVENT
5. Can you PREVENT a fall from occurring? If NO got to MINIMISE
6. Can you MINIMISE the distance and/or consequence of a fall

During an in-person assessment, personnel were asked to describe the task that they would be completing at each location and were asked to focus on what they thought the risks associated with working at height for the task would be. Once an initial understanding of potential risks were identified, personnel were asked to step through the task making use of the fall arrest equipment provided. Crash mats were placed around the work area to reduce the working at height risk of the assessment. Feedback was collected on the practicality of using the fall arrest equipment and any additional risks that its introduction may introduce.

What we experienced

Drawing from our experience of using the working at height guidance provided by the HSE the following observations are made on the utility and challenges associated with using the guidance.

Utility:

- Understanding how a task is completed is an important first step. It enables the practitioner to better understand what the task involves, who is involved and how it is to be completed.
- Better understanding of task conduct and completion allows for the practitioner to more easily assess whether potential working at height mitigations and protections will be appropriate and effective.
- Conducting a working at height assessment in-person at the location of the task with end-users allows for better understanding of the task, all that is involved and common practice amongst end-users.
- The three HSE provided questions were useful to guide discussion and ensure that all potential working at height mitigations are discussed.

Challenges:

- The guidance was not as expansive as first anticipated and left the practitioner questioning whether the working at height assessment was being approached correctly.
- HSE (2014) acknowledges that a sensible approach is needed when considering precautions for working at height as there are some low-risk instances where no particular precautions may be necessary. However, military vehicles are characterised by awkward footholds and are subject to a number of safety regulations. An approach that can allow for verification and validation of these instances would be more beneficial.
- In comparison to other HSE assessments, such as manual handling, it would have been beneficial to have more quantitative guidance to support the assessment. A particular gap was identifying what level of mitigation would be required for working positions of different heights.
- Military vehicles are characterised by protruding equipment and rough surfaces. There is little guidance to support the practitioner to understand and assess the effect falling from height. How an individual falls, and what they may hit on the way down, needs to be more clearly considered within the guidance.
- The threshold of when a mitigation is more dangerous than having no mitigation is not covered within the guidance. For example, for certain heights would the physical consequence from the pull of a harness when falling and the swing into the side of the vehicle be less dangerous than hitting the floor? Instances such as these fall outside of the low-risk instance where a sensible approach is accepted.
- No clear guidance is provided on what to do if none of the three decision points can be met. For example on operation, it may not be possible to avoid working at height, time criticality may mean that donning fall arrest equipment is not practical and there may be no infrastructure to allow the distance or consequences to be minimised.

Conclusion

It should be considered whether the working at height guidance can be expanded to provide a more comprehensive guide to practitioners, how guidance should include consideration of the effectiveness and risks associated with the mitigation itself and if a quantitative approach is appropriate.

References

Alizo, G. Sciarretta, J. D., Gibson, S., Muertos, K., Romano, A., Davis, J., & Pepe, A. (2017). Fall from heights: does height really matter? *European Journal of Trauma and Emergency Surgery. 44,* 411-416.

Army (2024a). Combat Vehicles: WARRIOR. Retrieved from https://www.army.mod.uk/learn-and-explore/equipment/combat-vehicles/warrior/

Army (2024b). Combat Vehicles: Stormer. Retrieved from https://www.army.mod.uk/learn-and-explore/equipment/combat-vehicles/stormer/

Defence Equipment & Support (n.d.). Challenger 2 and Challenger 3. Retrieved from https://des.mod.uk/what-we-do/army-procurement-support/challenger-2/

HSE (2014). Working at height: A brief guide. *Health and Safety Executive.* Retrieved from https://www.hse.gov.uk/pubns/indg401.htm.

HSE (2024). Work-related fatal injuries in Great Britain, 2024. *Health and Safety Executive*. Retrieved from https://www.hse.gov.uk/statistics/fatals.htm.

Complex task performance is predicted by integrative skill domain ability

Adrien Jouis, Marie Cahillane, Ken McNaught & Victoria Smy

Centre for Defence and Security Management and Informatics, Cranfield University; Defence Academy of the United Kingdom

SUMMARY

This paper presents a reanalysis of Lee et al’s (2012) skill acquisition data to test whether performance in a complex integrated task, based on fighter pilot attributes, is predicted by integrative skill domain ability. The findings show that the integrative skill domain ability is predictive of performance in a complex integrated task. In addition, more common individual difference measures like cognitive ability, working memory, attention control, and work sample testing, are either not predictive of performance or less predictive than the integrative skill domain ability.

KEYWORDS

Skill domain, performance, individual differences

Introduction

Psychological skill domains are a way to classify categories of cognitive processes underlying skills. Complex cognitive skill domains classify effortful cognitive processes that place significant demands on cognitive resources for the effective completion of tasks. One of these domains is the Integrative skill domain, which “*represents the ability of an individual to manage their attention to integrate and coordinate two or more concurrent psychological domains*” (Cahillane et al., 2022, p.10). This domain is required for complex integrated tasks where two or more components, underpinned by different psychological skill domains, must be performed concurrently rather than separately. High-risk tasks, representative of the application of the Integrative skill domain, are numerous, with piloting a fighter jet being a prominent example of a complex integrated task.

The potential effect of skill domain abilities on performance differs from the concept of skill transfer as, unlike task-specific skills, the psychological skill domains are not context specific. Instead, these domains can be considered “modes” of cognition, underpinned by different psychological processes and associated neural pathways (Lam et al., 2022; Strick et al., 2021, Worringer et al., 2019). As actionable modes of cognition they can be considered a form of individual difference. As such, measuring ability in relation to psychological skill domains could support the identification of operators with the required level of ability for effective performance during a given task.

In the context of the Integrative skill domain, as different complex integrated tasks share the same mode of cognition and neural areas (Ding et al., 2024, Peters et al., 2019), their performance is likely to be correlated as they rely on the same skill domain ability. This is especially true for Space Fortress performance; its relative difficulty and the integration of a wide-ranging variety of sub-tasks (Donchin, 1989) mean that performance in this task will likely impose a high cognitive load related to all the psychological and cognitive processes shared between complex integrated tasks. However, simpler, though still complex, integrated tasks, by not being as comprehensive, are more likely to only partially and unevenly load shared psychological processes and neural pathways (Leone et al., 2017). As they are likely to represent different load distributions, integrating multiple measures of simpler integrated tasks together should help with creating a measure that is more representative of individual performance across the whole shared neuro-cognitive pathway for complex integrated tasks. Such a composite measure should be predictive of very complex integrated task performance, as with Space Fortress, by representing integrative skill domain ability.

Various individual differences have already been used to predict performance in complex tasks, including multitasking paradigms (Ackerman, 1992; Draheim et al., 2022; Redick et al., 2016), but no research has been done considering complex integrated tasks as a category. Among individual differences, cognitive ability is one of the most studied constructs, with a robust effect on performance, though the effect seems to be smaller in recent literature (Berry, 2024). Cognitive ability, along with measures of memory and attention control, have found practical applications in personnel selection (Broach et al., 2019). However, cognitive ability measurements have been criticised as having an unequal effect for certain demographics, leading to an adverse impact within personnel selection (Burgoyne et al., 2021). Attention control and work sample tests have been proposed as alternative measurements (Burgoyne et al., 2021; Campion et al. 2019) although they have a limited association with performance (Roth et al., 2005).

Methods

Space Fortress is a video game representative of a complex integrated task. It is based on fighter pilot attributes, with motor demands, a need for visual monitoring and scanning, memory requirements (Donchin, 1989) and concurrent performance of skills underpinned by different psychological domains. This paper presents a reanalysis of Lee et al.'s (2012) data to examine whether a measure of Integrative skill domain ability predicts complex integrated task performance in Space Fortress, and how this compares with the predictive ability of existing individual difference measures.

In Lee et al.'s (2012) original experiment, 75 participants completed up to 15 sessions practicing Space Fortress in one of three conditions. The conditions were either a control, with three practice sessions, full emphasis training, with 15 sessions, or hybrid variable-priority training, with 15 sessions where the instructions changed. Each participant also completed three batteries of tests at the beginning, middle and end of the session series. Each test battery measured individual differences in cognitive ability, working memory, attention control, plus motor ability (single joystick task), a component skill of Space Fortress (piloting task, can be considered a work sample test), and performance in integrated complex tasks (radar monitoring task and dual-joystick task). For the reanalysis, we combined the two integrated complex task performance scores to obtain a composite measure representative of integrative skill domain ability. The regression analyses performed controlled for condition by session interactions. All the continuous variables (results of the test batteries and performance scores in Space Fortress) were Z-normalised. Data from each of the three batteries of test was regressed on the performance scores from the practice session which immediately followed. As the data had repeated measures and the residuals approximated a normal distribution, every regression was a robust linear mixed model ($k=1.345$, $s=10$). Every β reported is a standardised coefficient.

Results

Single skill domain multitasking, as represented by performance in an attention blink paradigm, mapped to a perceptual-visual domain, was not significantly predictive of performance[1] ($p = 0.749$, $t = 0.320$). Similarly, none of the individual difference measures of cognitive ability (Raven's

[1] Unless specified otherwise, the following regressions are controlled for the component skill.

progressive matrices: p=0.842, t=0.199), working memory (Sternberg memory task: p=0.079, t=1.764) or attention control (Flanker task: p=0.299, t=1.041) had a significant effect on performance. The component skill measure significantly predicts performance[2] (Piloting task: p=0.003, t=3.051; β =0.158, se=0.052), however the composite measure of Integrative skill domain ability has a stronger relationship with performance (integrative domain ability: p<0.001, t=5.311; β =0.221, se=0.042). This is reversed for the control sub score[3], the score where the component skill is most relevant, with the component skill explaining more variance than integrative skill domain ability (component skill: p<0.001, t=5.515; β =0.282, se= 0.051; integrative domain ability: p<0.001, t=3.494; β =0.154, se=0.044).

The individual measures that compose the integrative skill domain ability are less predictive of performance than the composite (radar monitoring: p=0.002, t=3.228; β =0.131, se= 0.041; dual joystick: p <0.001, t=3.930; β =0.167, se=0.043), validating the use of a combined measure, and they are uncorrelated (r(196)=0.020, p=0.783). A physical task representative of the Continuous psychomotor skill domain and fluency in the controls used by Space Fortress (i.e., use of a joystick), was not predictive of overall Space Fortress performance[4] (single joystick task: p=0.156, t=1.426).

Discussion

A significant effect of integrative skill domain ability on performance of the complex integrated task, Space Fortress, was found across training length and conditions. By contrast, individual differences in cognitive ability, working memory and attention control, which have previously been found to be significantly associated with overall complex integrated task performance (Redick et al., 2016) do not reach significance here, showing that they have more limited generalisability in explaining complex integrated task performance. Even a skill integral to Space Fortress performance and a form of work sample test, the component skill, and a skill that reflects the form of psychomotor control used in Space Fortress, have less influence than integrative skill domain ability on overall Space Fortress performance. This is a marked difference from the literature on skill transfer, where skill proximity to the task dictates the strength of the relationship between the two (Sala et al., 2019). A caveat to these results is that initial integrative skill domain ability does not predict long term performance, unlike the component skill.

Consequently, operator proficiency in, or potential capability for, complex integrated task performance could, in part, be trained and assessed more effectively by combining multiple simpler tasks that still require the integrative skill domain. This is because, in a standardised environment, one might assume that integrative skill domain ability is the primary limiting factor of performance, given the results from this study. If this holds true across applied settings, the design of complex integrated systems would benefit from regarding concurrent multi-task completion as the main challenge to satisfactory performance, rather than memory or attention. Designers of such systems should thereafter plan accordingly, limiting the operator multi-tasking requirements as much as possible and/or selecting people who are known to perform well in this domain. As the relationship between integrative skill domain ability and complex integrated task performance is maintained across training conditions, this ability and its underlying neuro-cognitive pathway are likely

[2] This regression is controlled for the composite score instead of the component skill.

[3] Subset of the Space Fortress score which tracks whether the player remains within the bounds of a hexagon visible on screen.

[4] This regression controls for the composite score instead of the component skill because of a lower Bayesian Information Criterion for this model.

affected by practice, suggesting targeted training interventions as an alternative to highly selective personnel recruitment.

Future research should confirm the assumption that this ability is the primary limiting factor of operator performance in applied environments. It should also aim to measure the benefits of an intervention, on individual integrative skill domain ability, for supporting complex integrated task acquisition compared to more traditional approaches. It would also be useful to examine whether the predictive power of integrative skill domain ability on performance generalises to other psychological skill domains. The findings have implications for personnel selection, transferability of experience and training.

Acknowledgements

We would like to thank Professor Walter R. Boot, Doctor HyunKyu Lee and Professor Arthur F. Kramer for giving us permission to use their data for a secondary analysis and helping us locate it. We would like to thank Doctor Kyle Harwell for providing the data and helping us understand its structure. We would like to thank Doctor Trevor Ringrose for his review of our statistical analysis. This research forms part of a PhD funded by Defence Science and Technology Laboratory.

Data supporting this study cannot currently be made available due to multiple third-party restrictions. Contact the lead author if you require access at adrien.jouis@cranfield.ac.uk

References

Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinants. *Journal of Applied Psychology*, *77*(5), 598–614. https://doi.org/10.1037/0021-9010.77.5.598

Berry, C. M. (2024). Personnel selection systems and diversity. *Current Opinion in Psychology*, *60*, 101905. https://doi.org/10.1016/j.copsyc.2024.101905

Broach, D., Schroeder, D. & Gildea, K. (2019) *Best Practices in Pilot Selection*. Office of Aerospace Medicine (DOT/FAA/AM-19/06). https://rosap.ntl.bts.gov/view/dot/57077

Burgoyne, A. P., Mashburn, C. A., & Engle, R. W. (2021). Reducing adverse impact in high-stakes testing. *Intelligence*, *87*, 101561. https://doi.org/10.1016/j.intell.2021.101561

Cahillane, M., MacLean, P. & Anderson, T. (2022). *Competence Retention Analysis Handbook*. Human and Social Sciences Research Capability (HSSRC) Technical Deliverable to UK MOD. O-HSSRC-1.004-029.

Campion, M. C., Campion, E. D., & Campion, M. A. (2019). Using practice employment tests to improve recruitment and personnel selection outcomes for organizations and job seekers. *Journal of Applied Psychology*, *104*(9), 1089–1102. https://doi.org/10.1037/apl0000401

Ding, Q., Ou, Z., Yao, S., Wu, C., Chen, J., Shen, J., Lan, Y., & Xu, G. (2024). Cortical activation and brain network efficiency during dual tasks: an fNIRS study. *NeuroImage*, 120545. https://doi.org/10.1016/j.neuroimage.2024.120545

Donchin, E. (1989). The learning strategies project. *Acta Psychologica*, *71*(1-3), 1–15. https://doi.org/10.1016/0001-6918(89)90002-4

Draheim, C., Pak, R., Draheim, A. A., & Engle, R. W. (2022). The role of attention control in complex real-world tasks. *Psychonomic Bulletin & Review*, *29*(4). https://doi.org/10.3758/s13423-021-02052-2

Lam, T. K., Vartanian, O., & Hollands, J. G. (2022). The brain under cognitive workload: Neural networks underlying multitasking performance in the multi-attribute task battery. *Neuropsychologia*, *174*, 108350. https://doi.org/10.1016/j.neuropsychologia.2022.108350

Lee, H., Boot, W. R., Basak, C., Voss, M. W., Prakash, R. S., Neider, M., Erickson, K. I., Simons, D. J., Fabiani, M., Gratton, G., Low, K. A., & Kramer, A. F. (2012). Performance gains from directed training do not transfer to untrained tasks. *Acta Psychologica*, *139*(1), 146–158. https://doi.org/10.1016/j.actpsy.2011.11.003

Leone, C., Feys, P., Moumdjian, L., D'Amico, E., Zappia, M., & Patti, F. (2017). Cognitive-motor dual-task interference: A systematic review of neural correlates. *Neuroscience & Biobehavioral Reviews*, *75*, 348–360. https://doi.org/10.1016/j.neubiorev.2017.01.010

Peters, S., Eng, J. J., Liu-Ambrose, T., Borich, M. R., Dao, E., Amanian, A., & Boyd, L. A. (2019). Brain activity associated with Dual-task performance of Ankle motor control during cognitive challenge. *Brain and Behavior*, *9*(8). https://doi.org/10.1002/brb3.1349

Redick, T. S., Shipstead, Z., Meier, M. E., Montroy, J. J., Hicks, K. L., Unsworth, N., Kane, M. J., Hambrick, D. Z., & Engle, R. W. (2016). Cognitive predictors of a common multitasking ability: Contributions from working memory, attention control, and fluid intelligence. *Journal of Experimental Psychology: General*, *145*(11), 1473–1492. https://doi.org/10.1037/xge0000219

Roth, P. L., Bobko, P., & Mcfarland, L. A. (2005). A Meta-Analysis Of Work Sample Test Validity: Updating And Integrating Some Classic Literature. *Personnel Psychology*, *58*(4), 1009–1037. https://doi.org/10.1111/j.1744-6570.2005.00714.x

Sala, G., Aksayli, N. D., Tatlidil, K. S., Tatsumi, T., Gondo, Y., & Gobet, F. (2019). Near and Far Transfer in Cognitive Training: A Second-Order Meta-Analysis. *Collabra: Psychology*, *5*(1), 18. https://doi.org/10.1525/collabra.203

Strick, P. L., Dum, R. P., & Rathelot, J.-A. (2021). The Cortical Motor Areas and the Emergence of Motor Skills: A Neuroanatomical Perspective. *Annual Review of Neuroscience*, *44*(1). https://doi.org/10.1146/annurev-neuro-070918-050216

Worringer, B., Langner, R., Koch, I., Eickhoff, S. B., Eickhoff, C. R., & Binkofski, F. C. (2019). Common and distinct neural correlates of dual-tasking and task-switching: a meta-analytic review and a neuro-cognitive processing model of human multitasking. *Brain Structure and Function*, *224*(5), 1845–1869. https://doi.org/10.1007/s00429-019-01870-4

Generation after Next HMI in Defence: What might the future look like?

Victoria Steane, Katie Shepherd & Erinn Sturgess

Research, Technology & Solution Innovation, Thales

SUMMARY

Future systems in defence are likely to incorporate increasingly levels of automation and Artificial Intelligence. With data proliferation representing a significant challenge, alternative visualisation and presentation technologies may be needed to better support operators in completing tasks. This paper aims to provide a view of the current state of the art and future trajectory of visualisation and presentation technologies. Centred on the task of tactical picture compilation, this paper describes the findings of a scoping review and technology scan aiming to identify potentially suitable approaches to support the visualisation and presentation of a tactical picture in Generation after Next (GaN) systems.

KEYWORDS

Defence, generation after next, human machine interface

Introduction

Trend research indicates that future defence systems will make use of increasing levels of automation and use of Artificial Intelligence. It is widely anticipated that such approaches will have multiple benefits at varying levels (i.e., individual operator, team, and organisation). In a general sense, automation and AI aims to improve operational effectiveness and efficiencies contributing to improved lethality and survivability. It is anticipated that automation and AI will improve operator performance by enhancing their decision-making (making 'smarter' and 'faster' decisions), reducing workload and improving situational awareness. Despite the potential of such approaches, we know from the academic literature that significant Human Factors challenges remain with these approaches (e.g., issues relating to complacency, inappropriate trust calibration, misuse). Further, from a Human Computer Interaction perspective, significant issues remain that relate to dealing with increasing volumes of data; fusing and synthesising data from different sensors (both local and remote); dealing with uncertain and ambiguous data; and achieving a common understanding. Whilst user interfaces have evolved to meet current requirements, this approach may not be suitable for future requirements (Fay et al., 2020). In short, new Human Machine Interfaces (HMIs) may be required to maintain effective performance (Fay et al., 2020).

Imagining what the future may look like can be a difficult endeavour. This is because Generation after Next (GaN) capabilities often refer to things that are not yet available or fully understood. For HMI specifically, this may mean incorporating components yet to be made possible. Generating future-orientated solutions and innovations requires some degree of Strategic Foresight (SF; Gordon et al., 2019). SF requires researchers to use a structured and systematic approach to explore potential future ways of operating. For example, identifying trends and scenario planning are widely used and common methods that are used to help develop foresight (e.g., Schwarz, 2008; Vallet et al., 2020). Literature reviews, expert panels, Delphi studies, and use of visual artefacts to imagine future scenarios are also often used (Popper, 2008; Kimbell, 2011; Mozuni & Jonas, 2017).

Method

As the initial phase of developing strategic foresight, this paper will describe the findings of a scoping review and technology scan which intended to explore the current state of the art and trajectory of visualisation and presentation technologies. The review was intentionally broad to explore alternative ways of supporting and enhancing operator performance but also identified a number of tools, techniques and technologies that demonstrate potential. To contextualise the analysis, a use case surrounding tactical picture compilation was utilised. Tactical picture compilation was chosen because it is a process that is conducted across all defence domains to gain situational awareness about the surrounding environment. Tactical picture compilation is based upon the integration of data from local and/or remote sensors/sources (both human and non-human) to form a visual representation about your surrounding environment. It incorporates information relating to objects (i.e., identification, classification), that are displayed within a geographical context (i.e., in relation to own position, or the position of others/key points of interest). Generally, a tactical picture will be formulated using all available data sources (e.g., visual, aural, vocal) and is a coordinated endeavour, involving many people and systems – but can involve uncertain or contradictory information, depending on the sensors and situation. Further, in pursuit of multi-domain operations, there will need to be a shift from platform-centric approaches to domain-centric approaches. This will mean capitalising on fused data and multi-static processing. There is increasing emphasis on Multi-Sensor Data Fusion (MSDF) given its potential to improve decision-making and reduce cognitive workload (Kessler & White, 2017).

In order to understand specific challenge areas associated with tactical picture compilation, an empathy map was created with support from an ex-operator within the maritime domain. Empathy Mapping is a powerful tool within the solution innovation space as it enables you to build empathy and resonate with the intended user group allowing designers to consider how they may be better supported in fulfilling their role/task. They were first asked to consider how work is currently done and then, in light of potential future scenarios, consider how this may change. A number of challenge areas were identified using this approach (Figure 1).

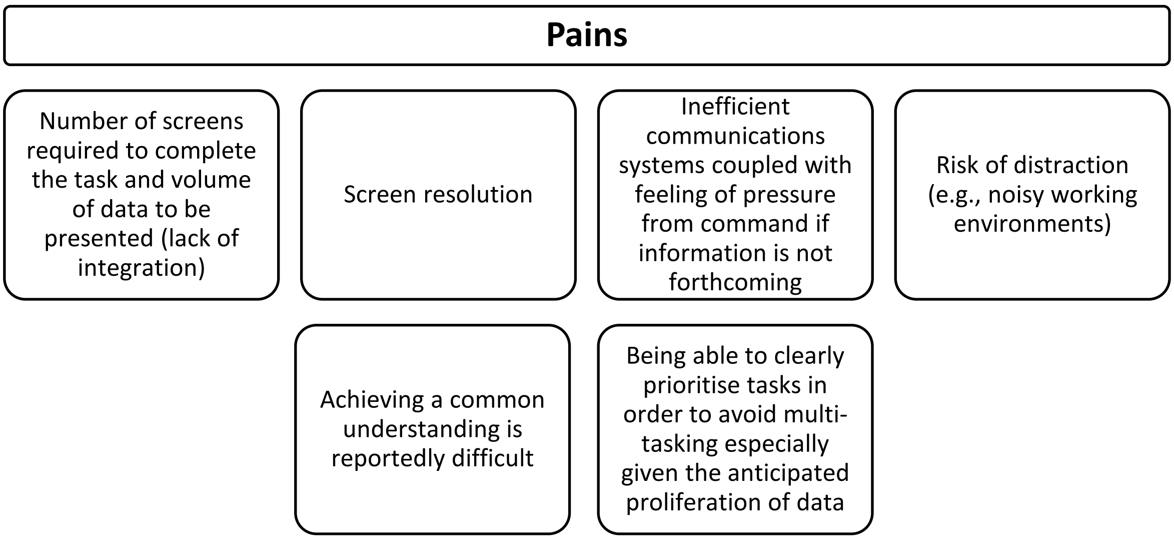


Figure 1: Challenge areas identified as relevant to tactical picture compilation

The suitability of alternative HMI approaches to this type of task were then explored further using a Red, Amber, Green (RAG) assessment – a robust approach offering Subject Matter Experts (SMEs) a way of identifying solutions that demonstrate most potential. The criteria used for assessment were broadly based on desirability, feasibility and viability metrics to allow for a holistic view to be taken (Figure 2), whilst also recognising the challenge areas identified above. For example, applicability to the task of tactical picture compilation and dissemination; integration with the system, training implications and safety.

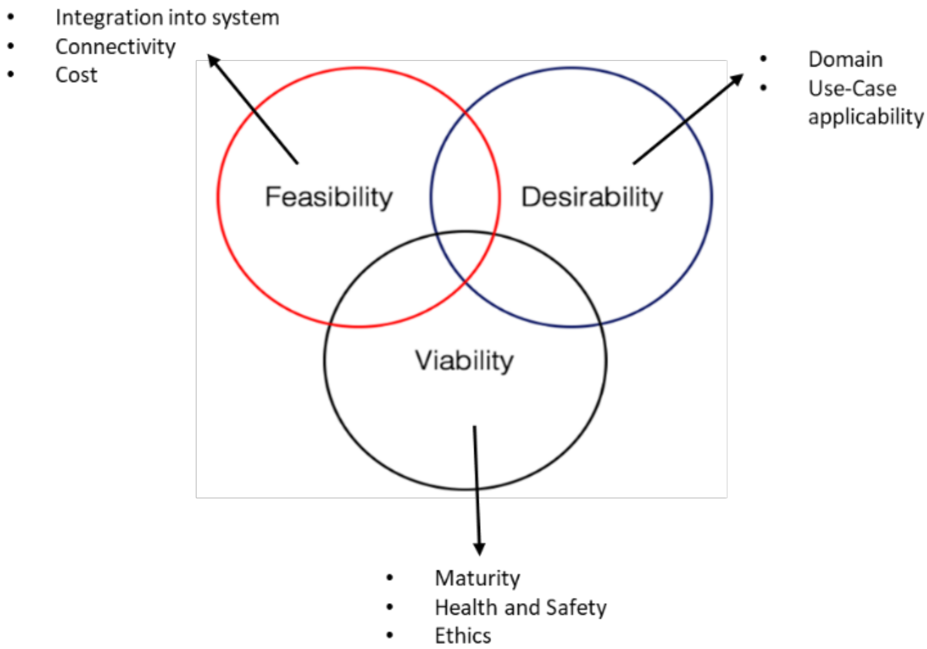


Figure 2: Connected nature of key assessment concepts

Findings

The RAG assessment identified a large volume of visualisation, presentation and decision support technologies as being of interest for future systems in which tactical picture compilation features. These included but were not limited to:

- 3D audio displays;
- Artificial Intelligence;
- Automation;
- Decision Support Systems; and
- Mixed Reality.

Despite the large volume of technologies identified, understandably not all were rated as being suitable for the task of tactical picture complication. For example, whilst 3D audio displays may be applicable to other roles within command teams, aural data is not used within target motion analysis. Some of the most promising technologies identified through the RAG assessment include:

Three dimensional (3D) displays

3D displays vary in size and the format of the display, yet provide the opportunity to present 360 degrees of information. Depending on the size and format of the display, 3D visual displays may be used to encourage co-location and collaborative working between team members and assist in the planning of operations. It is anticipated that the initial training burden for 3D visual displays will be moderate, as it represents a significant step change away from traditional 2D displays.

Whilst the detail on the technical development of 3D displays is beyond scope of the current work, it is important to acknowledge that images are presented to the eye using temporal or spatial interlacing. Temporal interlacing is prone to temporal artefacts (e.g., flicker, distortions in perceived depth) whilst spatial interlacing can be prone to poorer spatial resolution (Banks et al., 2016). However, in general, there is some evidence to suggest that 3D visual displays can improve operators situational awareness (e.g., Lager et al., 2019), user experience (e.g., Pitts et al., 2015) and improve overall system safety (e.g., Lager et al., 2019).

3D visual displays have been widely used within the civilian space for some time (e.g., mechanical design in the automotive and aviation sectors, medical imagery and architecture; Rousseau, 1994). In the context of tactical picture compilation, Rousseau (1994) argued that 3D visual displays would present informational components more intuitively, particularly for above and below water scenarios. Anti-Submarine Warfare displays, for example, could include the provision of high-resolution, computer generated imagery, pertaining to the environment (e.g., a representation of active, passive, location and environmental sensor information) and its relevance to the local bottom topography and water properties (Rousseau, 1994).

Augmented Reality (AR)

AR has been heralded as a technology that can be used to improve the ability of individuals to perceive information and performance in tasks leading to enhanced global awareness (e.g., Kim & Dey, 2016). Within the maritime sector, the use of AR technologies aim to support and improve operator situational awareness (Grundmann et al., 2022). However, it represents a relatively new technology for maritime operations meaning that the effects on operator performance are not yet widely understood (Van den Oever et al., 2023). Further, applications of AR within the maritime sector include ship navigation, construction, maintenance, inspection and training so more research is needed to fully establish the suitability of AR technologies to the task of tactical picture compilation. In the short term at least, AR may be suitable for training new operators (Patterson et al., 2010) or be used as a mechanism to overlay important information directly into an operator's field of vision supporting the interpretation of real-time information. According to Lackey et al., (2014), simulation based training that emulates the real world is more likely to facilitate the transfer of learning to operational contexts. The utility of AR will be in part determined by the specific use case under scrutiny.

Holographic Displays

Unlike traditional 3D visual interfaces, holographic displays provide operators with the capability to move around and view different angles of the same image. It is anticipated that holographic displays will enable a more intuitive visualisation of the tactical picture as they enable assets and contacts to be visualised in their 3D positions and motion vectors. Within the marketplace, there are alternative forms of holographic displays. "Sandbox" implementations enable multi-person collaboration whilst "monitor" implementations are somewhat smaller but still enable up to three people to collaborate together (e.g. Urban Photonic Sandtable Display by the Defense Advanced Research Projects Agency); and has also been implied to aid situational awareness (Fay et al., 2019). In the context of command and control, both 3D and holographic HMI solutions may provide a platform in which a more open dialogue between team members may be achieved as they offer a means to view and plan in real-time, increasing the speed of decision-making.

Alternative visual HMI displays

Alternative HMI displays represent a significant step-change away from the traditional means of presenting data to tactical picture compilers. Touch foils and projection screens, for example, offer a way to enlarge the display area and permit greater collaboration amongst team members within control room environments. Other types of displays (e.g., mid-air displays and rollable displays) permit greater levels of flexibility in terms of the location in which tactical picture compilation may take place. Stanney et al., (2004) argued that graphics are better than text or auditory instructions when you are trying to communicate spatial information as they produce better comprehension of complex tasks. However, interfaces that allow for more active engagement or direct manipulation, as those identified above) are thought to lead to better comprehension of information, supporting users in dealing with and comprehending 'uncertainty' (Newton et al., 2017). Given that tactical picture compilation often involves handling ambiguous data, greater levels of embodied interaction with visualisations may support situational awareness and operator cognitive processing.

Discussion

Imagining the future can be challenging – particularly with a 2060 time horizon. However, technology scans provide the opportunity to explore the current state of the art and trajectory of visualisation and presentation technologies moving forward. This type of strategic foresight contributes to our understanding of future display technologies and their potential use in supporting and enhancing the work completed by tactical picture compilers in GaN systems remembering that "speculation unfettered by display constraints leads to some intriguing possibilities" Rousseau (1994, p.30). Whilst current technologies may not yet provide the level of sophistication required to fully realise its potential, the literature base points to many advantages of alternative HMI tools, techniques and technologies.

Moving forward, we must remain mindful that vision is the predominant sense used to convey information to humans yet there are many other senses that can be exploited to transfer information. For example, multi-modality interactions are likely to offer greater levels of enhancement than visual interventions alone in situations whereby operators are exposed to huge volumes of data. Multiple Resource Theory (MRT; Wickens, 2002) may offer a good foundation to guide the design of GaN HMI, particularly in situations whereby an operator is required to perform multiple tasks simultaneously. This is because MRT suggests that distributing tasks across different sensory modalities can reduce dual task interference, which should, in turn lead to more efficient information processing and better task performance (Wickens, 2002). Stanney et al., (2004) published a comprehensive set of guidance relating to optimal senses to convey different information types. Whilst this research is over twenty years old, the guidance still appears to be valid.

Future research will continue to use strategic foresight approaches in combination with Design Thinking approaches to further explore GaN HMI within the context of tactical picture compilation across a number of defence sectors (i.e., land, air, sea). Using Design Thinking provides a platform to co-create innovations that connect the needs of the intended end users with technical solutions. In this sense, it an alternative way of thinking about and approaching problems in a user-centred way. Inviting end users to be directly involved within ideation processes will provide further insight into "how" they may be better supported, "what" they might need and "when" they need it. This data can contribute significantly to the design of specifications for future systems.

References

Banks, M. S., Hoffman, D. M., Kim, J., & Wetzstein, G. (2016). 3D Displays. *Annual Review of Vision Science*, *2*(1), 397-435.

Fay, D., Stanton, N. A., & Roberts, A. P. J. (2019). All at sea with user interfaces: From evolutionary to ecological design for submarine combat systems. Theoretical Issues in Ergonomics Science, 20(5), 632-658.

Fay, D., Stanton, N. A., & Roberts, A. P. J. (2020). Designing future submarine control room HMIs. Contemporary Ergonomics and Human Factors 2020.

Gordon, A. V., Rohrbeck, R., & Schwarz, J. O. (2019). Escaping the 'faster horses' trap: Bridging strategic foresight and design-based innovation. Technology Innovation Management Review, 9(8), 30-42.

Grundmann, R., Hochgeschurz, S., Hohnrath, P., & Ujanki, A. (2022). White paper: Increasing maritime situational awareness by augmented reality solutions. Fraunhofer Center for Maritime Logistics and Services CML.

Kessler, O., & White, F. (2017). Data fusion perspectives and its role in information processing. In Handbook of Multisensor Data Fusion (pp. 35-64). CRC Press.

Kim, S., & Dey, A. K. (2016). Augmenting human senses to improve the user experience in cars: Applying augmented reality and haptics approaches to reduce cognitive distances. Multimedia Tools and Applications, 75, 9587-9607.

Kimbell, L. (2011). Rethinking design thinking: Part I. Design and Culture, 3(3), 285-306.

Lackey, S. J., Salcedo, J. N., Matthews, G., & Maxwell, D. B. (2014). Virtual world room clearing: A study in training effectiveness In Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC). Orlando, FL. 14045

Lager, M., & Topp, E. A. (2019). Remote supervision of an autonomous surface vehicle using virtual reality. IFAC-PapersOnLine, 52(8), 387-392.

Meshram, V. V., Patil, K., Meshram, V. A., & Shu, F. C. (2019). An astute assistive device for mobility and object recognition for visually impaired people. IEEE Transactions on Human-Machine Systems, 49(5), 449-460.

Mozuni, M., & Jonas, W. (2017). An introduction to the morphological Delphi Method for design: A tool for future-oriented design research. Journal of Design, Economics, and Innovation, 3(4), 303-318.

Newton, O. B., Fiore, S. M., & LaViola Jr, J. J. (2017). An external cognition framework for visualizing uncertainty in support of situation awareness. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 61, No. 1, pp. 1198-1202). Sage CA: Los Angeles, CA: SAGE Publications.

Patterson, R. E., Pierce, B. J., Bell, H. H., & Klein, G. (2010). Implicit learning, tacit knowledge, expertise development, and naturalistic decision-making. Journal of Cognitive Engineering and Decision-making, 4(4), 289-303.

Pitts, M. J., Hasedžić, E., Skrypchuk, L., Attridge, A., & Williams, M. (2015). Adding depth: establishing 3D display fundamentals for automotive applications (No. 2015-01-0147). SAE Technical Paper.

Popper, R. (2008). How are foresight methods selected? Foresight, 10(6), 62-89.

Rousseau, D. (1994). 3-D Displays and controls for sonar operators. Virtual Reality Systems, 1(2), 28-32.

Schwarz, J. O. (2008). Assessing the future of futures studies in management. Futures, 40(3), 237-246.

Stanney, K., Samman, S., Reeves, L., Hale, K., Buff, W., Bowers, C., & Lackey, S. (2004). A paradigm shift in interactive computing: Deriving multimodal design principles from behavioral and neurological foundations. *International Journal of Human-Computer Interaction*, *17*(2), 229-257.

Vallet, F., Puchinger, J., Millonig, A., Lamé, G., & Nicolaï, I. (2020). Tangible futures: Combining scenario thinking and personas-A pilot study on urban mobility. Futures, 117, 102513.

Van den Oever, F., Fjeld, M., & Sætrevik, B. (2023). A Systematic Literature Review of Augmented Reality for Maritime Collaboration. International Journal of Human–Computer Interaction, 1-16.
Wickens, C. D. (2002). Multiple resources and performance prediction. Theoretical Issues in Ergonomics Science, 3(2), 159-177.

Human-Centred Assessment of Human Augmentation Technologies

Victoria Cutler, Alison Clerici & Eleanor Cox

QinetiQ Ltd, UK

SUMMARY

Human Augmentation (HA) technologies have been identified as a key future technology to enhance human performance, which could be of benefit in a range of contexts, including defence and security. However, there are a wide range of HA technologies, and limited methods available to evaluate the benefits and risks associated with their use. This project tested an approach to evaluating HA technologies that involved a modified version of the Ministry of Defence (MOD) Early Human Factors Analysis (EHFA). The HA EHFA was tested by applying it to the use of telexistence for use in battlefield medical care. The HA EHFA was successful in being able to identify the operational benefits, capability vulnerabilities, and ethical considerations associated with the technology. It is recommended that the HA EHFA be used to evaluate HA technologies for use in a defence and security context.

KEYWORDS

Human Augmentation, Early Human Factors Analysis

Introduction

Human Augmentation (HA) technology has been defined by the Ministry of Defence (MOD) as "*the application of science and technologies to temporarily or permanently improve human performance*" (MOD, 2021). HA technologies seek to improve performance of the user to, or beyond, their biological potential. As this technology continues to develop, there is a need to understand the potential benefits of HA for defence and security, whilst considering the safety, wellbeing, and performance of the user. This understanding is necessary to allow constrained budgets to be invested in the most promising technologies, to de-risk the use of HA technologies, and maximise its intended impact.

Work undertaken by QinetiQ on behalf of the Defence Science and Technology Laboratory (Dstl) has explored the use of the MOD Early Human Factors Analysis (EHFA) Methodology (MOD, 2016) to assess HA technologies. EHFA is typically used in the early stage of a procurement process, and is a structured and systematic method to identify the human-related benefits and risks of a technology.

EHFA could provide a means of characterising the benefits and risks of HA technology, as it ensures human capabilities, needs and limitations are taken into account across all contexts of use. EHFA is also a broad approach, covering all elements of human system integration including the context, equipment, organisation, operation, maintenance, training, health hazards, and system safety. This breadth would be beneficial for HA, as HA covers a wide range of technology types which have a range of different impacts for users.

The initial work concluded that a modified EHFA could be used to assess the use of HA technologies in a specified military context, but that further testing and piloting was required to refine and validate the process (Clerici, 2023).

A follow on task was undertaken, also funded by Dstl, which aimed to refine and test the HA EHFA in the context of an ongoing research programme. The task set out to review and validate the HA EHFA, and develop an HA EHFA Methodology Guide. Accordingly, the project was structured into three phases as shown in Figure 1.

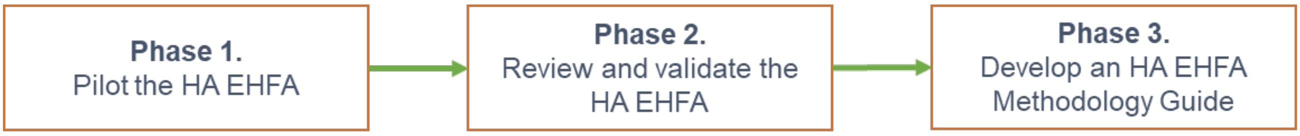


Figure 1: Project structure

This paper presents the outcomes from this project, with a focus on the development of the HA EHFA methodology.

Phase 1. Pilot the HA EHFA

The HA EHFA was piloted on the use of telexistence for battlefield medical care. Telexistence is the use of robotic avatars, operated remotely by a human in an immersive environment, to perform a task. Telexistence augments the human by allowing them to feel present in an environment other than the one they are currently in, allowing them to operate in that environment as if they were there. This use case was chosen to align with a Dstl research programme investigating the potential uses of telexistence. The pilot HA EHFA aimed to identify if telexistence could be applied to achieve competitive advantage in the context of battlefield medical care.

The first stage of the HA EHFA involved a review of telexistence documentation and interviews with two stakeholders. One stakeholder was an expert in telexistence the other in battlefield medical treatment. Two stakeholders was adequate for the pilot nature of the EHFA, and the emerging nature of the technology. A series of workshops were then used to establish, review, and score considerations associated with implementing telexistence in battlefield medical care. The considerations were documented in a Human Factors Integration Risks Assumptions, Issues, Dependencies, Opportunities, Ethical Concerns (HFI RAIDO-E) Register.

The HFI RAIDO-E Register for the use of telexistence for battlefield medical care contained 53 HFI considerations, of which there were 37 risks, 1 assumption, 6 dependencies, and 9 opportunities. No issues were identified. The HFI Considerations covered a range of technical areas. The majority related to equipment, process, and user integration. However, there were also HFI Considerations related to safety, social and organisational factors, training, and user characteristics.

Risks and opportunities were rated for their probability and impact, and this was used to generate an overall Probability-Impact Score between 1 and 6, using the scoring table in the MOD EHFA Methodology Guide (MOD, 2016). Assumptions and dependencies were not given a Probability-Impact Score, as they do not have a quantitative probability or impact. There was a spread of Probability-Impact Scores, as shown in Figure 2, indicating that the process was able to identify a subset of risks and opportunities as the most important for effective integration of the user. Examples of risks and opportunities with a high Probability Impact Score are:

- Reduced risk of harm to the Medic (opportunity);
- Trust (risk); and
- Casualty acceptance (risk).

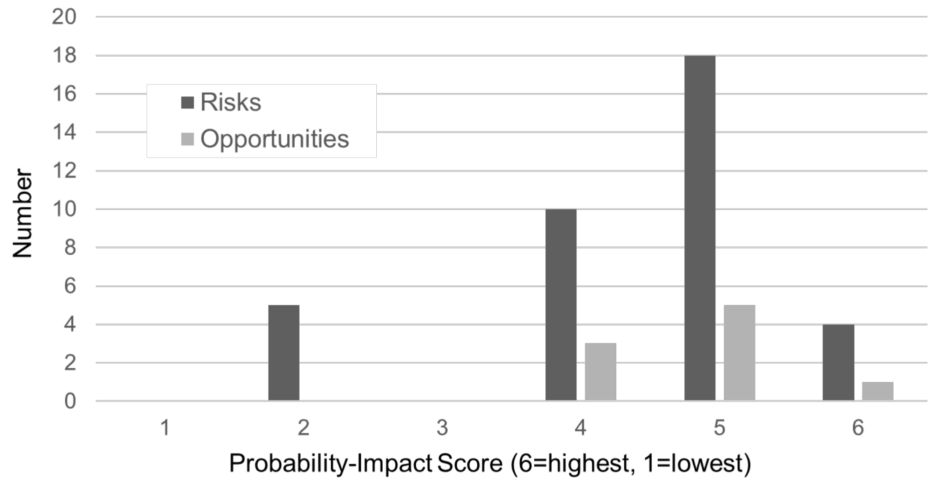


Figure 2: Number of risks and opportunities with each probability impact score

Using the HFI RAIDO-E Register, a Capability Evaluation was carried out. The Capability Evaluation involved determining an ethical stance and summarising the operational benefits, capability vulnerabilities, and HA technology development status. A summary of areas highlighted is shown in Table 1.

Table 1: Summary of the Capability Evaluation

	Operational benefits	**Capability vulnerabilities**	**Ethical issues**
Definition	Competitive advantage relative to the UK's existing capability and/or adversarial capability, by application of the HA technology in the specific context	Any weakness/issue that might lead to an enemy's enhanced ability to destroy, degrade, disrupt or deny a UK capability/advantage	Ethical concerns associated with the HFI Considerations identified.
Number identified in pilot EHFA	9	6	8
Examples from pilot EHFA	• Reduction in risk to medics • Remote monitoring • Equipment carriage	• Casualty acceptance • Disruption to communications • Loss of data	• Standard of medical care delivered • Importance of human contact • Medical confidentiality

Following the Capability Evaluation a decision point was reached where the team must determine, based on the ethical stance, operational benefits, capability vulnerabilities and technology development status, whether to continue with the final stages of the HA EHFA process. For the pilot EHFA, it was decided that the capabilities could outweigh the vulnerabilities but only with further development of the technology, and that further assessment was required in relation to the ethical issues. This meant that the response plans were developed with the goal of reaching the point at which a telexistence system could be procured for use in battlefield medical care.

Response plans were generated for each HFI consideration, often comprising of multiple actions. The actions were grouped into related tasks and an overall plan of activities generated showing the order in which the tasks would be performed, and how the tasks were linked. This sequence of HF activities is shown in Figure 3.

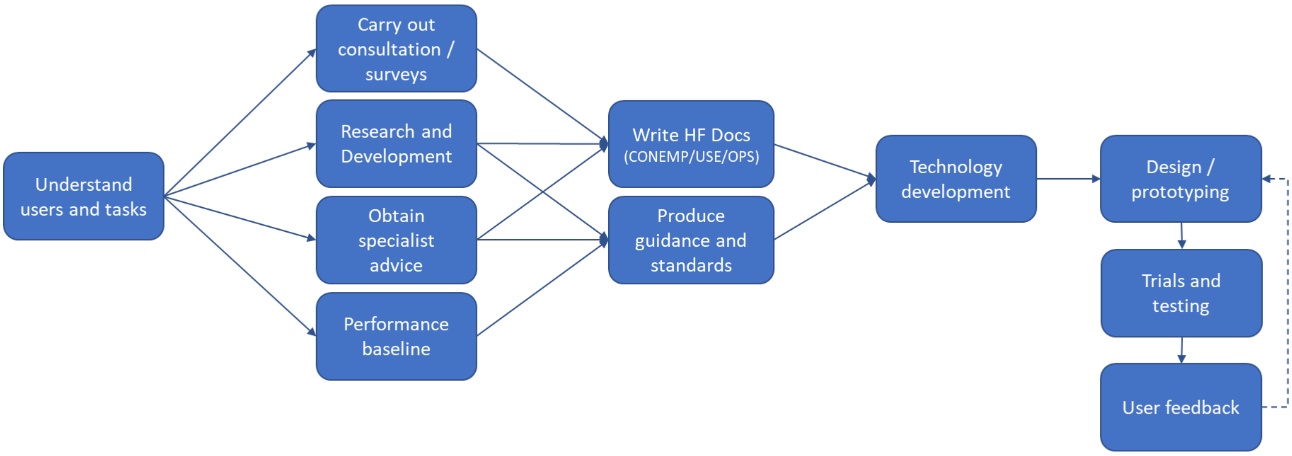


Figure 3: Sequence of HF activities

At the end of the pilot, the HA EHFA was documented in a report and a separate HFI RAIDO-E Register. The HA EHFA Report described the process applied and the HFI Considerations, presented the Capability Evaluation Summary, and detailed the response plans. The HFI RAIDO-E Register provided a full set of data on each of the individual HFI Considerations, and was designed to be a living document that could be used to manage the HFI Considerations through the future development of the HA technology.

Phase 2. Review and Validate the HA EHFA

Completing the pilot HA EHFA on using telexistence for battlefield medical care gave the team insight into the effectiveness of the process as originally devised. The team held an internal review to draw out these insights which:

- Evaluated the HA EHFA process, based on the experience of applying it to the potential use of telexistence for battlefield medical care;
- Reflected on the wider applicability of issues identified with the HA EHFA process; and,
- Identified further refinements to the HA EHFA.

The review concluded that the HA EHFA could be applied successfully in real world applications. The EHFA outputs were usable, addressed the aim of the EHFA, and were suitable to inform decision makers. A particularly valuable output of the HA EHFA was its ability to identify discrepancies between Technology Readiness Level (TRL) and Human Readiness Level (HRL).

A critique of the process was that the HA EHFA outputs were at a high level, but this was considered appropriate to the stage of telexistence development at which it was conducted. The most challenging part of applying the HA EHFA method was the ethical assessment process, and a number of updates to this aspect were recommended.

The results of the HA EHFA and internal team review were presented to external stakeholders in a validation workshop. The aims of the workshop were to:

- Gain initial validation of whether the proposed HA EHFA methodology is fit for purpose;

- Expose the work with communities who are likely to exploit and/or benefit from the modified process; and,
- De-risk and inform specifications for developing the HA EHFA Methodology Guide.

The workshop was attended by representatives from Dstl, Defence Equipment and Support (DE&S), MOD Front Line Commands (FLC), and ethics SMEs. The consensus at the workshop was that the HA EHFA provided a useful means of assessing HA technologies in a defence or security context. It was also highlighted that the methodology would benefit from further modifications to maximise its utility, particularly around ethics.

Phase 3. Develop a HA EHFA Methodology Guide

The findings from the internal review of the HA EHFA and the validation workshop were used to update the HA EHFA method and develop an HA EHFA Methodology Guide. The HA EHFA Methodology Guide was prepared in the same format as the existing MOD EHFA Methodology Guide to maximise its potential exploitation.

The HA EHFA Methodology Guide outlined five stages to the HA EHFA as shown in Figure 4. These five stages are the same as the MOD EHFA Methodology, but there are refinements to each stage that support the aims of the HA EHFA, and tailor the approach to the nature of HA technologies.

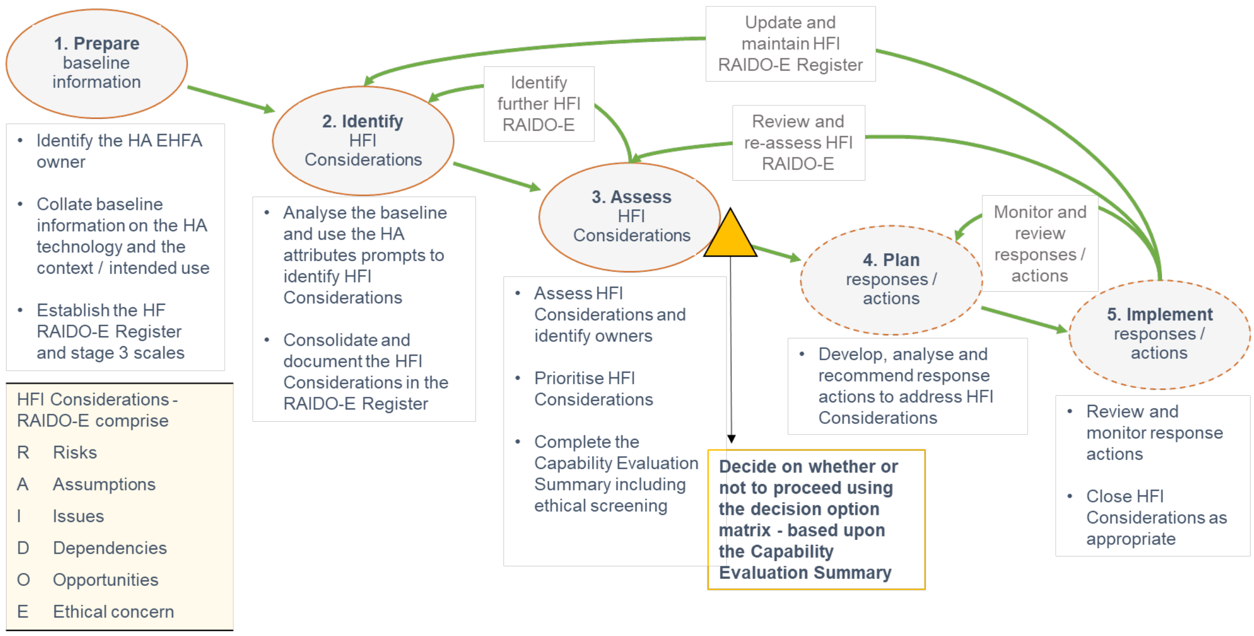


Figure 4: HA EHFA Process

The HA EHFA Methodology Guide also included resources to support the conduct of the HA EHFA including prompts to be used to gather information relevant to HA technologies, scales to assess the HFI Considerations, a table highlighting HA attributes for consideration during the EHFA, and a template for conducting the Capability Evaluation.

Conclusions

The pilot study of telexistence for battlefield medical care showed that the HA EHFA Methodology was effective in identifying ethical, safety, and other HFI considerations. The HA EHFA outputs were usable and suitable to inform decision makers. While the outputs from the pilot study were at a high level, this was appropriate for the telexistence project for which it was conducted. The HA EHFA was also effective at producing a focused and manageable set of response plans, by setting a realistic goal state to be achieved by those actions.

It was concluded that the HA EHFA Methodology is suitable for use in assessing the use of HA technologies, at Research and Development (R&D) stage, in a defence context. The review and validation phase identified areas where modification of the HA EHFA as piloted were required to ensure that the outputs were representative and usable for its intended purpose. These changes were made and integrated into the HA EHFA Methodology Guide. The HA EHFA Methodology Guide can be obtained from the Dstl Human Augmentation team.

It is recommended that specific and more extensive ethical guidance be developed to support the evaluation and use of HA technologies in defence and security contexts. To support the exploitability of this work further, it would also be beneficial that the HA EHFA is integrated into the MOD EHFA Methodology Guide.

References

Clerici, A. (2023). A process to assess the use of human augmentation technologies in defence. Paper presented at Ergonomics and Human Factors 2023.

Human Factors and Ergonomics Society (2021). Human Readiness Level Scale in the System Development Process. ANSI/HFES 400-2021.

Ministry of Defence (2016). Early Human Factors Analysis (EHFA) Methodology Guide. Technical Note Issue 1.2

Ministry of Defence (2021). Human Augmentation – The Dawn of a New Paradigm. A Strategic Implications Project.

On Using AcciMap to Support Judgements of Risk During System Development

Mike Tainsh

BAE Systems Maritime

SUMMARY

AcciMap is a tool developed for the investigation of system failures during past events. This empirical study on the judgements of 12 specialists investigated judgements of potential failure associated with system development using the AcciMap technique. The system development studied was a radar centre for the defence of London in the late 1930s. The results showed that system development experts make consistent judgements on some AcciMap characteristics but not others including external influences and some job designs.

KEYWORDS

System development, risk assessment, AcciMap

General Introduction and Aim

System development teams are keen to mitigate risks during the development process, so that the product has well founded characteristics when "in service". One ergonomics/human factors technique in support of this aim is Early Human Factors Analysis (EHFA) to identify risks with a view to mitigation. This use of EHFA has been central to the application of ergonomics/human factors as required by the UK Ministry of Defence (2016) over at least the past two decades. The judgements of risk involve estimates of both likelihood and impact of design characteristics. However, little is known about how well these judgements are made during system development.

Ergonomics specialists have known for many years that people can be poor in making judgements of risk (Cohen, 1960) and that these judgements can be easily influenced by a variety of circumstances. Further, these judgements can be influenced during group decision making (Kogan and Wallach 1967) when group influences can increase the likelihood of making risky judgments. However, in a system development context, we have no empirical evidence to indicate how good or bad, systems development specialists may be in making judgements that underpin the assessment of risk. Further, many of these judgements are made in a team context.

This investigation set out to survey the judgements of system development specialists with a view to gaining a better understanding of their strengths and weaknesses when makings judgments of likelihood and impact. These judgements underpin the assessment of risk.

One technique employed to understand the characteristics of past failures is "AcciMap". However, in this case, the AcciMap technique was used to look forward into the future deployment of a system rather than backward to understand failings as exemplified by Waterson (2023). Further the technique needed to be linked to a task involving judgements that might be undertaken during system development where there is a need to avoid failure. Hence, it was recognised from the start that a characterisation scheme was required that supported the study of judgements about the future.

Waterson (2023) summarised his approach to AcciMap using five categories which support the characterisation of failures:

- external influencers,
- organisational issues,
- workplace constraints,
- physical constraints and
- outcomes.

These categories may be considered in terms of a set of layered descriptions that can be traced from one to another and be functionally decomposed from the highest to lowest levels. The linkages between the layers give traceability within the development process.

Firstly, the external issues in system development are the system goal and associated conditions which must be specifically defined to ensure that its functional decomposition yields specific criteria at each layer. Secondly the organisational layer provides a description of the operation/business which determines the functions which the system must perform. Thirdly, the workplace must be described as they determine the conditions and constraints which are a result of the equipment characteristics and the tasks which must be performed to ensure effective and safe operation. The final two layers describe the tasks for the users within the system and their individual and personnel characteristics.

The Technique

The AcciMap categories are given in the left hand column of Table 1. These categories have to be mapped onto system characteristics which are relevant to system development.

The system characteristics relevant to the user are given in Table 1. The mapping is made to characteristics used to describe a User System Architecture (USA) as described in earlier studies (Tainsh, 2018).

The USA forms part of the total system architecture for the capability. However, to understand the risks within the design, it is necessary to understand the possible causes of failures that may occur over the anticipated lifetime of the system. This is true for equipment as well as organisation development, or combinations of the two. Once the possible causes of failure are identified then mitigation can be put into position.

An understanding of the extent of the risk is important to support studies of their mitigation as these will help identify the likelihood and impact of adverse events occurring. The case for mitigation will depend on judgements of these two factors.

Table 1 uses the concept of Layers. The USA is understood to be open to functional decomposition where each layer is available for decomposition into the layer below. The exact process of decomposition will depend on the specific layers under consideration. The Layer number is a shorthand means of reference.

Table 1: AcciMap summary as applied to USA within investigation (Waterson et al 2017)

Layer number	AcciMap Categories	USA Layer Name	The features of Biggin Hill air defence capability at RAF Biggin Hill
1	External influences – System goal and capability	User goal, scenario/ context and constraints	Defend London Air Defence Area (LADA).
2	Organisational issues	Business/operational description	Conduct coordinated Air Defence from Biggin Hill to counter the threat through the use of a

			three layered system of assets (an outer layer of guns, middle layer fighters and an inner layer of balloons and barrages).
3	Workplace constraints, Physical constraints	A technical description of the system's equipment including user aspect	Use of radar, plotting tables and secure telephony with specialist military teams.
4	Team /task issues	Users' roles and team organisation	Use of standalone radars by specialist RAF teams to plot aircraft movements and intelligence, to feed information to asset controllers.
5	Work personnel issues	Individual User's characteristics and tasks	Air defence picture compilers/plotters Asset managers

The Capability of Air Defence of London in the late 1930s

The example of capability development used here is Air Defence at RAF Command at Biggin Hill (Zimmerman, 20024). This development was undertaken in highest secrecy during the late 1930s.

The air defence capability for London was achieved, in part, through a radar system with the user architectural characteristics shown in Table 1. This formed part of the total system architecture for the capability which in turn linked into the overall defence capability of the UK. Clearly the radar system is part of a larger defence system including air and other assets.

The development intention is to have requirements that result in performance effectiveness that satisfied criteria of RAOAL (Risks At Operationally Acceptable Levels). These criteria address the lifecycle of the system.

The RAOAL assessment criterion is associated with an appropriate risk assessment matrix. In this context "highly unlikely" may be quantified as not expected to change within the next year but maybe within two-to three years, "unlikely" as potentially could happen within the year but not expected, "likely" means that an impact could be expected within the first two quarters of the year. Guidance on the meaning of the impact assessment was provided to the participants for each judgement.

The impacts are quantified dependent on the characteristics of the system. The evaluation of the combination of impact a likelihood is given in Figure 1 which uses a set of categories ranging from trivial, tolerable, moderate, substantial up to intolerable. Clearly the latter two categories are the most severe and to be avoided. All categories of risk should be addressed proportionately i.e. with project effort matching risk severity as part of the mitigation process.

		Potential degree of impact/performance loss		
		Systems experiences slight performance loss	Systems experiences minor loss of performance	Systems involves major loss of performance
Likelihood of loss occurring	Highly unlikely	Trivial	Tolerable	Moderate
	Unlikely	Tolerable	Moderate	Substantial
	Likely	Moderate	Substantial	Intolerable

Figure 1: Applicable Risk Matrix with outcome evaluations

The Study

A study was conducted with participants who were all employees of BAES with substantial experience of system development. It was not a homogeneous group - all had varied backgrounds as users, engineers, managers and ergonomics specialists.

Each participant was given a 14 page proforma which described the development of radar as taken from Zimmerman (2004) and described each of the USA layers with a description of the impacts as shown in Table 2.

Method

The investigation was conducted by the author. It included describing the radar development process which was detailed within the proformas to each participant. The proformas included a figure for recording the risk decisions. It was also explained that the task of making the judgements should be carried out under calm conditions that enabled each participant to take full account of all the information that was provided in the proformas. The participants carried out this task away from the workplace.

The means of categorising likelihood was explained along with the specification of categories of impact for each layer.

It was explained that a single cell of the risk matrix had to be selected as this was the risk that was to be used for developing the radar system. Mitigation would depend on the evaluation associated with it.

The participants completed the task independently but were able to ask for advice to clarify information and ensure their understanding. The data from all the twelve participants was tabulated in histograms and each histogram was compared to a random distribution using a chi-squared technique to assess whether there was consistency in the judgements.

The results

The results are given in full in Table 3. In summary:

- The judgements on external influence (Layer 1) showed no consistency.
- The judgements on the design of the operational system (Layer 2) showed a high degree of consistency with a moderate risk.
- The judgements on the design of the technical system (Layer 3) showed a high degree of consistency with a moderate risk.
- The judgements on the design of the team organisation (Layer 4.1) showed a significant degree of consistency with a moderate risk.
- The judgements on the design of the Duty Officer's job (Layer 4.2) showed no consistency.
- The judgements on the design of the Plotters' jobs (Layer 4.3) showed a significant degree of consistency with a substantial risk.
- The judgements on recruitment and training (Layer 5) showed a high degree of consistency with intolerable risk.

Conclusion

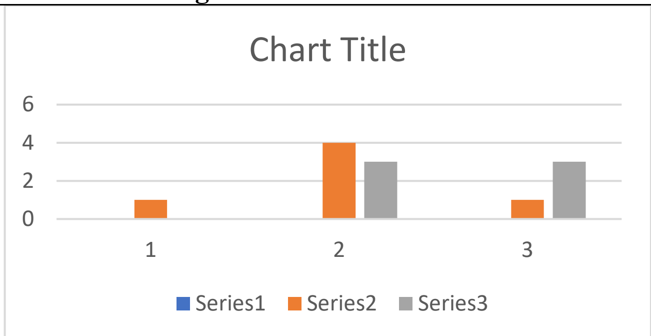
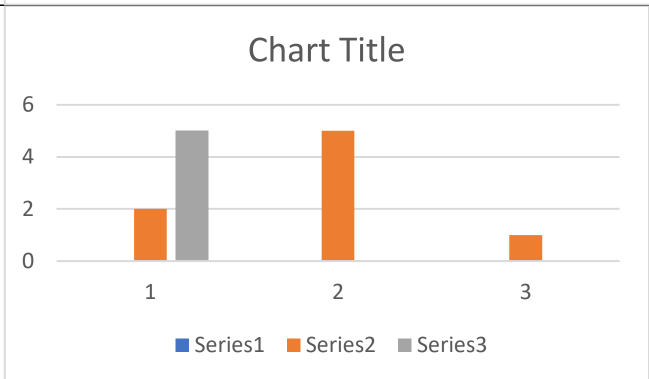
It was found that there was consistency in 5 out of the 7 sets of judgements.

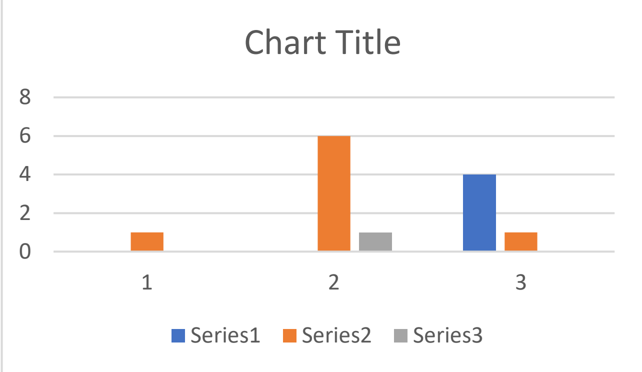
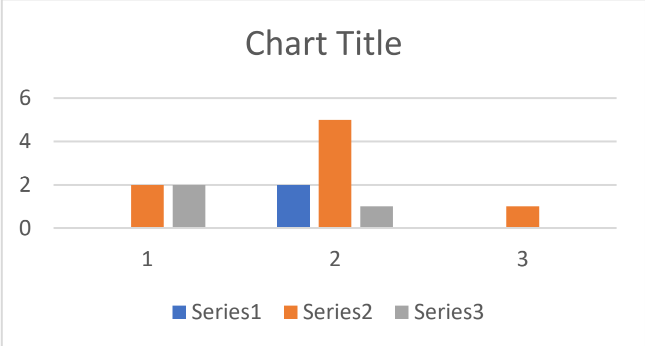
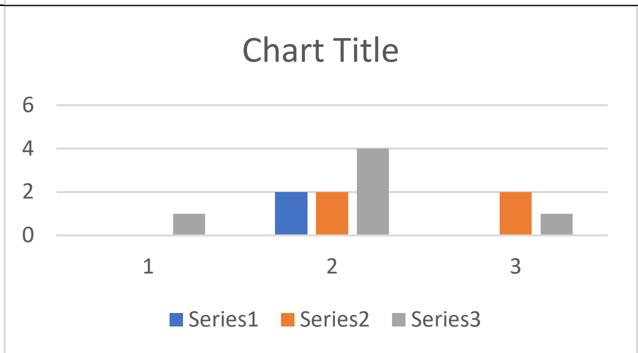
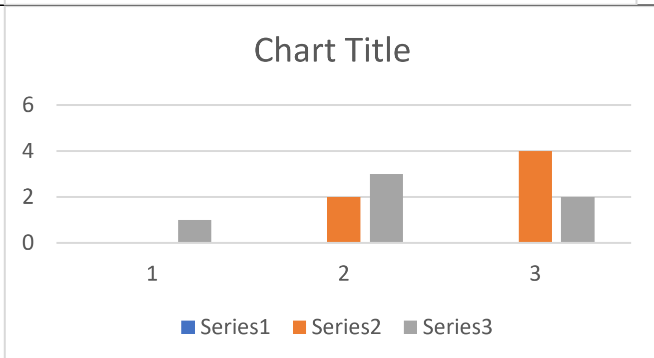
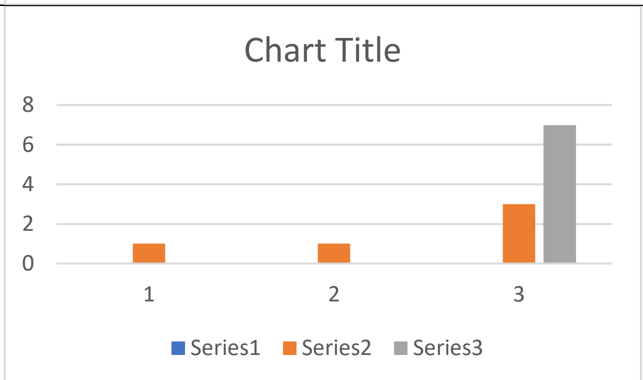
Table 2: Categories of Impact

1 User goal, scenario	Slight	A result of a change in German formations or aircraft weapons.
	Minor	A result of a change in German number of aircraft or means of protection.
	Major	A new class of weapons deployed or a major switch of resources within theatre
2 Organisat-ional/ Operational description	Slight	A result of poor operational practice within the operational system.
	Minor	A result of low capability in limited parts of the system.
	Major	A result of low capability in substantial portions of the system.

3 The system's equipment	Slight	A result of equipment sometimes being unreliable or failing.
	Minor	A result of parts of the equipment set failing but the remainder maintaining operational capability.
	Major	A result of substantial equipment failures.
4.1 Users' roles/ structure	Slight	A result of slight problems of communication or flow of information.
	Minor	A result of a problems of information flow and authorisations in a constrained part of the Command organisation.
	Major	A substantial failure in a pat of the organisation that put the operation at risk of complete failure,
4.2 Users' roles/tasks/ Duty Officer	Slight	A result of the Duty Officer experiencing difficulties in carrying out the task.
	Minor	A result of the Duty Officer experiencing acknowledged performance failures.
	Major	A result of the Duty Officer failing in an important part of his duties.
4.3 Users' roles/tasks/ Plotters	Slight	A result of the Plotters experiencing difficulties in carrying out the task.
	Minor	A result of the Plotters experiencing acknowledged performance failures.
	Major	A result of the Plotters failing in an important part of his duties.
5 Users' recruitment and training	Slight	A result of the Plotters experiencing difficulties in carrying out the task which could be linked to recruitment and training.
	Minor	A result of the Plotters experiencing acknowledged performance failures. which could be linked to recruitment and training?
	Major	A result of the Plotters failing in an important part of their duties which could be linked to recruitment and training.

Table 3: The results showing the raw data and statistical significance

Layer number	**Layer Name**	**Histogram of frequency of participants' judgements (total 12) for each risk matrix. Series 1, Series 2 and Series 3 show impact assessments in increasing magnitude and blue, orange and gray show decreasing likelihoods.**	**Chi-squared prob-ability**
1	User goal, scenario/ context and constraints.		Not significantly different from random
2	Organisational/operational description		Significance greater than 0.001

3	A technical description of the system's equipment including user aspect		Significance greater than 0.001
4.1	Users' roles and team organisation		Significance greater than 0.01
4.2	The Duty Officer		Not significantly different from random
4.3	The Plotters		Significance greater than 0.01
5	Individual User's characteristics and tasks – recruitment and training.		Significance greater than 0.001

Zimmerman does not state how the radar development team approached risk mitigation in the 1930s, but it is clear that an understanding of the external threat may have been difficult to take into account. It appears that some judgments were made consistently by the team but the system design depends on a judgement of external influences and here there was uncertainty. It suggests that the linking of development and external influences needs to be close or else the wrong system will be developed. Fortunately for the UK, in the 1930s, for radar, some good judgements were made by the UK.

AcciMap has helped characterise the USA development, and it appears that BAES employees can made consistent judgements on some development characteristics but have difficulty on two categories. The judgements on external influences are important because of their implications throughout the development. This indicates that engineering judgements as part of EHFA can be consistent but judgements of external events and job design may be less certain.

References

Cohen, J. (1960). Chance, skill, and luck: The psychology of guessing and gambling. Penguin.

Kogan, N and Wallach, M A (1967) Risky-shift phenomenon in small decision-making groups: A test of the information-exchange hypothesis. Journal of Experimental Social Psychology, Volume 3, Issue 1, January 1967, pp75-84

Tainsh, M (2018) Do our complex systems meet requirements? An example from Naval ergonomics. Contemporary Ergonomics & Human Factors 2018, CIEHF

UK Defence Standard 00-251 (2016) Human Factors Integration for Defence Standards

Waterson P (2023) Masterclass on AcciMap. CIEHF Conference 2023.

Zimmerman, D. (2004) Information and the Air Defence Revolution, 1917–40, Journal of Strategic Studies, 27:2, 370-394

Visualisations and storytelling in defence: Establishing requirements for a service offering

Siobhan E Merriman, Robert Becker, Freya Leith, Kolby D Pistak & Steven Wilson

AtkinsRéalis, UK

SUMMARY

Visualisations and storytelling help to communicate and facilitate the understanding of complex information, thereby supporting problem solving and optimal decision making. Interviews were conducted with 19 stakeholders to discover the needs and requirements for a new Visualisation and Storytelling Service (VSS) within a Defence context. A Thematic Analysis produced six key themes and 18 recommendations for the VSS. This paper will explain the discovery methodology used, and the themes and recommendations generated for the future initialisation of the VSS.

KEYWORDS

Visualisations, Storytelling, Defence, Interviews, Thematic Analysis, Requirements

Introduction

In the current data-driven and fast-paced environment, the ability to assess and understand complex information quickly is vital for informed decision making. Visualisations (e.g. graphics, videos) and storytelling help make information easier to comprehend, quicker to process and easier to transfer to long-term memory, thereby aiding performance (Kosara & Mackinlay, 2013). It is for these reasons that visualisations are used across numerous domains, including health care (e.g. Crackerjack Visual Thinking, n.d.), education (e.g. Bobek & Tversky, 2016) and transportation (e.g. Hansson, 2020). In this project, the Human Factors project team was asked to support the development of a Visualisation and Storytelling Service (VSS) for Defence where programme teams can procure a visualisation to help them communicate complicated ideas quickly. The intention was for the VSS to create videos and graphics to educate people on technical topics, communicate complicated problems/solutions, and explain the complicated background behind decisions to support optimal decision making and enhance performance of the UK Defence system of systems and those working within it.

Method

This project followed the Government Digital Service framework stages of Discovery, Alpha, Beta and Live. The project team participated in the Discovery phase, over a four-week period. A user-centred approach was taken to discover the needs and requirements for the VSS. The project team conducted interviews with potential stakeholders (users) to understand the demand, scale, size and requirements for the VSS. The findings were then used to develop recommendations for the future initialisation of the VSS. During the interviews, the project team was asked to gather key information regarding the volume of demand, complexity of outputs required, output classification, budget and any other considerations which may impact demand on the VSS.

Participants

A preliminary user research phase (which is outside the scope of this paper) generated ten personas to represent potential stakeholders (users) of the new VSS. Nineteen stakeholders were invited to attend a one-hour interview over Microsoft Teams, each representing one of these ten personas.

Semi-structured interview process

Twelve core questions to be asked in all interviews were generated. These were designed to produce insights regarding how the stakeholders currently work, how they could benefit from the new VSS, what their key needs for the service were and any other factors that could influence the effectiveness of the VSS. The questions covered their desired visualisation content, format of current evidence summaries, information classification, volume, branding and key target audiences. Where interviews finished early, additional questions were asked to gain further and more in-depth understanding of stakeholder needs (e.g. other types of visualisations, current visualisation challenges). Each interview included two project team members: a primary facilitator and an assistant. The facilitator led the interview and asked the majority of the questions. The assistant was invited to ask questions at various points, to gain clarity and ensure all core questions were covered. Both team members took detailed notes during the interviews, and these were reviewed and collated following each interview to accurately capture stakeholders' responses.

Data analysis

A Thematic Analysis, following the guidance in Braun and Clarke (2006), was used to analyse the interview data. The analysis produced six key themes; these represented the themes that were mentioned most frequently across the interviews and answered the key questions around volume, complexity, classification, and budget.

Results and key findings

The Thematic Analysis revealed six main themes in stakeholder responses:

- Tailored content – Stakeholders emphasised the need for content that accommodates diverse learning preferences, problem spaces, information needs and time constraints.
- Holistic view of the problem space – Stakeholders valued the ability to see the 'big picture' and zoom out to understand the broader system, identify gaps and avoid duplicate efforts (i.e. not procuring a capability that has already been acquired elsewhere).
- Communicate at the right level of detail – Content should strike a balance between complexity and simplicity, using accessible language (e.g. consistent use of terms, not full of jargon or overly technical language) and format (e.g. not a lengthy report).
- Demand for visualisation expertise – Many stakeholders create their content internally with little visualisation expertise. This leads to slower creation times and variable quality, in part due to the use of non-specialist applications like Microsoft Office tools.
- Practical considerations for producing visualisations – Information classification and content volume varied based on project/programme demands and needs (e.g. quarterly, all stages of the project lifecycle, one per project, multiple per project).
- Frequency of updates – Stakeholders indicated different levels of demand for content updates, including once in a project lifetime, annual updates, quarterly updates, once every project milestone and monthly updates.

These themes were then used to propose 18 recommendations for the future development of the VSS. The next phase of the project involved designing and initialising the VSS Alpha Service, incorporating the recommendations.

References

Bobek, E., & Tversky, B. (2016). Creating visual explanations improves learning. Cognitive Research: Principles and Implications, 1(27), 1-14.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77-101.

Crackerjack Visual Thinking. (n.d.). Great feedback from Manchester University NHS Foundation Trust about the 'Transforming Care for the Future' rich picture created by Crackerjack Visual Thinking #engage #communicate. Retrieved from CRACKERJACK VISUAL THINKING: https://crackerjackvisualthinking.com/2018/10/02/great-feedback-from-manchester-university-nhs-foundation-trust-about-the-transforming-care-for-the-future-rich-picture-created-by-crackerjack-visual-thinking-engage-communicate/

Hansson, L. (2020). Visual representation in urban transport planning: Where have all the cars gone? Transportation Research Part A: Policy and Practice, 133, 1-11.

Kosara, R., & Mackinlay, J. (2013). Storytelling: The Next Step for Visualization. Computer, 46(5), 44-50.

X-ing the Gap

Barry Peter Kirby, Amanda Georgina Kirby & Jim Wilson

K Sharp Ltd

SUMMARY

The GAP X project aimed to develop a lightweight gap-crossing solution for Armoured Fighting Vehicles (AFVs) to improves mobility on the battlefield. This paper summarises the project's three key phases, spanning 16 months, with a particular focus on the role and value of Human Factors Integration (HFI).

KEYWORDS

Human Factors Integration, Defence, Bridges, Armour

Project Overview

The GAP X project developed a short gap-crossing capability for Armoured Fighting Vehicles (AFVs). The system, developed by Beardsell & Son with Human Factors (HF) support from K Sharp, is a short-span temporary bridge to improve battlefield mobility of vehicles like the Boxer Infantry Fighting Vehicle (IFV). This project ran from March 2023 to July 2024, with three phases of HF involvement, managed through a tailored Human Factors Integration (HFI) approach. The GAP-X system is aimed to be operated under high-stress, time-sensitive conditions by military personnel. The integration of HFI at this pre-concept/early prototype stage ensured that the system was designed with the user in mind, focusing on ease of operation, safety, and reliability in adverse environments.

Approach

The HFI aspects had three main interaction phases as described in Figure 1. Each phase focused on a review of progress and developing feedback for the client to build into their next evolution of work.

Phase 1: Initial Concept Review (March 2023)

- A desktop review of the conceptual design, with an assessment of the system's potential Human Factors issues.
- The primary goal was to assess how operators of the system would transport, deploy, cross and recover the GAP X in a battlefield context as well as maintain it through life.

Phase 2: Prototype Field Trials (August 2023)

- A two-day field event was conducted to simulate real-world gap-crossing scenarios.
- The system's deployment, operation, and recovery were observed and recorded in conditions that aimed to reflect real world operational challenges.
- This enabled a systematic review of the whole operation that was then categorised by sub-system and operational phase

Phase 3: Final Field Trials and Enhanced Prototype (July 2024)

- A second set of trials on a refined prototype under more challenging conditions.
- The trial was conducted in a procedurally similar approach to the first set of trials, however the vehicle utilised was a closer representation of the target platform as well as an increase in weight.

Figure 1: Methodological Approach

Main Findings

The main findings can be split down into the three phases, which demonstrates how they evolved over the life of the project.

Phase 1: Initial Concept Review: This review identified 47 issues that were fed back to the client to enable them to consider as they moved into their prototype development stage.

Phase 2: Prototype Field Trials: This review identified 44 new issues and enabled a refresh of the 47 issues from the phase 1. Some examples of issues found were:

- *Hitching and Disconnection*: The NATO hitch proved difficult to disengage under certain conditions, leading to delays and potential safety risks.
- *Controller Accessibility*: The handheld controller was awkwardly positioned and difficult to access, increasing the time required for deployment and posing risks in high-stress environments.

Phase 3: Final Field Trials and Enhancements: Improvements were made to the GAP X system. The final field trials in July 2024 demonstrated changes including the following:

- *Improved Hitch Mechanism*: A quick-release mechanism was added to the hitch, making it easier and safer to disengage the system under challenging conditions.
- *Controller Repositioning*: The controller was moved to a more accessible location, and the connection process was streamlined, significantly reducing the time required for setup.
- *Wider Deployment Surface*: The bridge surface was widened to increase stability during crossings, especially for heavier vehicles like the Boxer IFV.

New Findings from Phase 3: 129 potential issues were identified through observation of the trial activity and the previous analysis. Examples of the issues highlighted included:

- *Roles and Responsibilities*: The operator roles and communication channels need to be clearly defined to ensure everyone knows what is going on.
- *Operator Training*: While the system's technical improvements reduced complexity, proper operator training was still critical to ensuring smooth deployment and recovery.

Key Contributions of Human Factors Integration (HFI)

Throughout the GAP X project, HF was provided input to the system's evolution, ensuring that user safety, operational efficiency, and environmental adaptability were considered at each stage.

1. *Early Identification of Critical Issues*: HF assessments provided insights into potential operational challenges. These early observations allowed for proactive design changes, reducing the likelihood of costly rework later in the development cycle.
2. *Design Improvements Based on User Feedback*: The phased approach allowed the team to incorporate feedback at each stage of development. Observations from the live trials in Phase 2 directly informed the modifications made in Phase 3.
3. *Operational Reliability and Environmental Adaptability*: HFI recommendations ensured that the system could adapt to a variety of environmental conditions, from muddy terrain to uneven surfaces.

Conclusion

The GAP X project demonstrates the critical importance of early HFI involvement in the pre-concept phases of military systems. Through continuous assessment and user feedback, the project was able to refine its design, addressing key usability and safety concerns that emerged during trials. On the face of it, the number of issues rose during this project, however the team reviewed the previously identified issues at every stage and retired those that had been demonstrably closed and then added those issues identified on the new version. Whilst the product was demonstrably more mature, it was still at the Concept/Prototype stage, and the ideal time to provide the feedback because it was the most cost-effective stage of procurement to implement solutions.

Augmented reality in earthquake rescue: impact on workload and decision making

Weixuan Li, Glyn Lawson, Setia Hermawati & Kyle Harrington

Human Factors Research Group, University of Nottingham

SUMMARY

This study investigates the application of augmented reality to enhance rescue efficiency, alleviate workload, and improve decision-making in earthquake rescue. The findings demonstrate that AR-based solutions surpass conventional approaches in task completion duration, map utilisation, and decision-making simplicity, underscoring AR's capacity to improve human factors issues in disaster contexts. Future studies should include professional rescue teams as well as more complex simulations to test AR's effectiveness in real-world emergency situations.

KEYWORDS

Augmented Reality, Earthquake rescue, Performance, Workload, Decision making.

Introduction

The utilisation of Augmented Reality (AR) technology in earthquake rescue is a significant advancement in disaster management. Earthquakes frequently produce catastrophic damage, needing an immediate and accurate reaction to find and rescue persons trapped within collapsed structures (Macintyre et al., 2006). Conventional rescue operations are often obstructed by restricted sight, insufficient situational awareness, and communication difficulties. Augmented Reality provides a revolutionary solution by augmenting real-time visualisation, strengthening spatial awareness, and promoting communication among rescue teams. Augmented Reality (AR) overlays digital information onto the physical world, allowing rescuers to effectively "see through" debris, map unreachable zones, efficiently coordinate operations Hu et al. (2022) and more generally improve rescue personnel's perceptual abilities (Demirkan & Duzgun, 2020; Xu et al., 2024). AR technology features such as head-mounted displays and interactive interfaces, have the potential to revolutionise the way search and rescue efforts are done in difficult and complicated places.

Several studies have emphasised the use of augmented reality in improving situational awareness, and navigation among rescue workers. For example, a system that combines ground-penetrating radar (GPR) and AR, such as the framework proposed by Hu et al. (2022), can support the visualisation of voids beneath disaster debris in 3D. This method was shown to dramatically enhance the efficiency and safety of searching for survivors trapped inside collapsed structures (Hu et al., 2022). Xu et al. (2024) investigated the application of multi-user AR platforms to facilitate team-based search and rescue operations. Their findings suggest that during search and rescue (SAR) missions, combining an exocentric navigation perspective with synchronised spatial data can improve collaboration efficiency. Chalimas and Mania (2023) proved the usefulness of a cross-device AR system that combines thermal imaging and GPS tracking to aid in firefighting and rescue operations. Finally, the importance of sophisticated communication technology in AR systems is particularly worth noting, as it can significantly improve the intelligence exchange capabilities of rescue teams. Wang et al. (2023) presented a 5G-enabled augmented reality architecture for exchanging real-time information during emergencies. This technology enables seamless data sharing between rescue teams and medical facilities, resulting in rapid and informed decision-making.

Based on previous research, AR has increased rescue efficiency, safety, and teamwork by improving perception and information aggregation. However, there is a lack of research on using AR in large-scale disasters like earthquakes. Therefore, this study focuses on earthquake rescue and explores whether AR can impact human factors issues in earthquake rescue.

Methodology

This study investigated the benefits of AR in earthquake rescue scenarios using a simulated virtual environment. A virtual earthquake environment was created with Unity 3D, and participants performed tasks in two modes: an AR-based system with holographic image map and 3D marking system and a traditional system similar to paper/digital maps. The study used a within-subject design, with participants receiving instruction on judging the level of building damage before the experiment. NASA-TLX was used to record performance measures such as task accuracy, completion time, and workload evaluation. Following the experiment, key decision analysis on the marking process was performed using video recordings of the experimental method and participant interviews to assess the decision-making process. Data from Unity logs and participant itineraries aided both quantitative and qualitative analyses of rescue performance. Research questions: Can AR offer advantages in workload, work memory and decision making in emergency rescue?

Participants

This study publicly recruited 20 ordinary participants through the internet. All participants conducted the experiment individually. Each participant was required to participate in both traditional rescue mode and AR-based earthquake rescue experiments. The order of participating in these two experiments was randomised to control for order effects. The basic information of the participants is as follows:

- Age: 18 - 65
- Gender: Both male and female
- Background: No professional rescue experience

Materials and Equipment

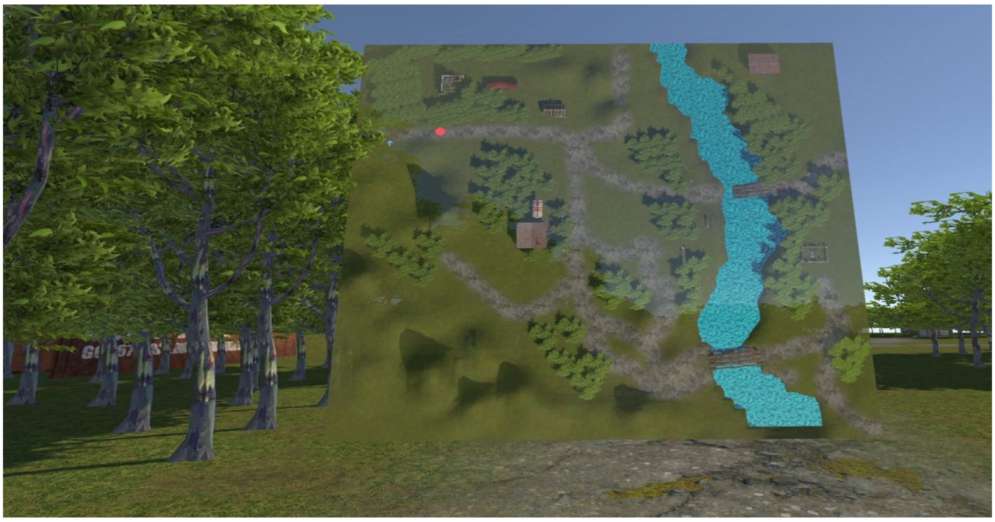


Figure 1: Mountain Village Earthquake Scene Completed Based on Unity 3D

- Virtual Environment: Mountain village earthquake scenes developed using Unity 3D, including multiple houses, collapsed buildings, rivers, bridges, and forests (Figure 1).
- Quest VR headset: Through the Quest VR headset, participants can immerse themselves in virtual earthquake rescue scenes and interact with the virtual environment.
- NASA-TLX questionnaire: Used to evaluate participants' perceived workload after the experiment.
- Data recording tool: Real time recording and uploading of participants' task completion time, walking path, and decision-making process data to the researcher's Raspberry Pi server.
- Interview tools: Conduct interviews using CDM Pro and experimental videos to collect participants' subjective feelings and key decision points in the decision-making process.

Design

This study uses a within-subjects design, with the independent variable being the rescue mode type, which includes:

- **Traditional Rescue Mode**: Use traditional maps (Figure 2) (digital maps) and marking on maps.
- **AR-based Rescue Mode**: Use AR map (Figure 3) (a translucent map that can be displayed in front of you at any time) and use AR technology to mark directly on the building.

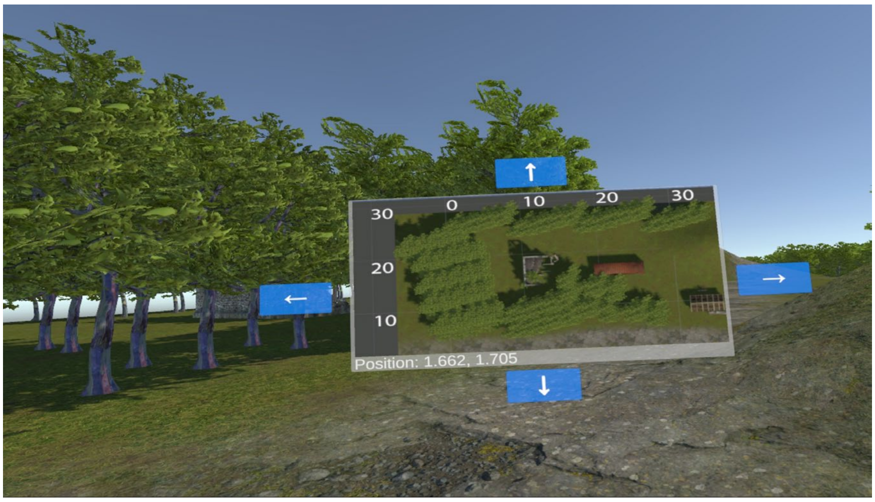


Figure 2: Traditional map

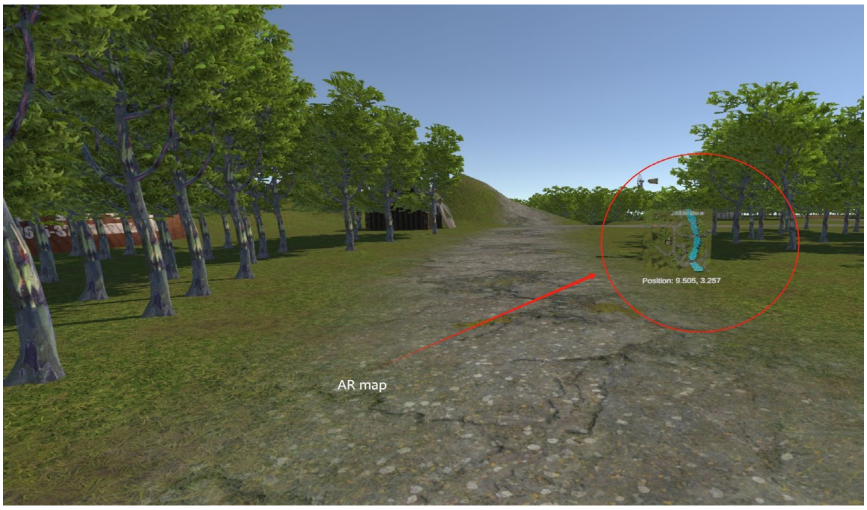


Figure 3: AR map

Reliability and Validity

To ensure the reliability and validity of the research results, the following measures were taken in this study:

- Internal reliability: By adopting a within-subject design, each participant engages in both rescue modes, thereby minimising the impact of individual differences.
- Randomization: The order of experimental conditions is randomly assigned to control for order effects.
- Standardisation procedures: The experimental process and task instructions are strictly standardised to ensure that all participants conduct the experiment under the same conditions.
- Multiple data sources: Combine quantitative and qualitative data to provide a comprehensive analytical perspective and enhance the validity of the results.
- Reliability of questionnaire: Utilise the validated NASA-TLX questionnaire to ensure the reliability and consistency of workload assessment.
- Interview reliability: Interviews were conducted using the CDM-Prob method to ensure systematic and consistent data collection.

This study uses these parameters to assure the trustworthiness of experimental data and the effectiveness of conclusions, thereby providing a sound scientific foundation for the use of augmented reality technology in earthquake rescue.

Result

The performance metrics show that the AR system outperforms traditional systems in every aspect of rescue operations. Statistical analysis using a paired-samples t-test shows that the AR system significantly reduces task completion time (t(df)=8.710, p=0.000, Cohen's d=1.948), indicating a significant improvement in efficiency. This effect is both statistically significant and practically meaningful, as Cohen's d value indicates a very large effect size. Furthermore, the AR system contributes to a moderate but significant increase in the number of correct labels for damaged buildings (t(df)=-2.096, p=0.050, Cohen's d=-0.469). Although this result is marginally significant, more research is required to confirm its impact.

In addition, the AR system decreased the frequency with which rescue personnel opened maps (t(df)=8.669, p=0.000, Cohen's d=1.938), making the task more fluid and requiring less cognitive effort. It also reduced the amount of time spent viewing maps (t(df)=6.737, p=0.000, Cohen's d=1.506), allowing for better access to critical information. The reduction in map usage time resulted in faster decision-making and task completion.

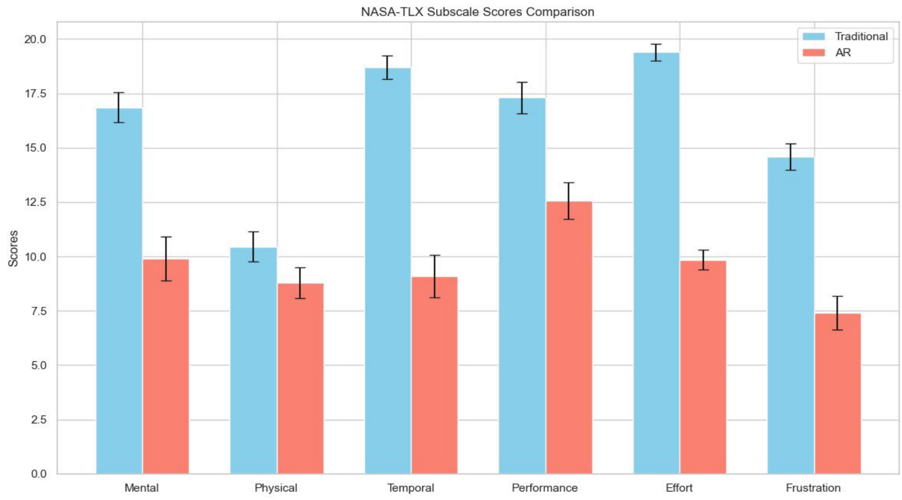


Figure 4: Comparative Bar Charts of NASA-TLX Subscale Scores for Traditional and AR Rescue Systems

Furthermore, the AR system significantly reduced the total walking distance required to complete the rescue mission (t(df)=3.215, p=0.005, Cohen's d=0.719), indicating that it is effective at optimising rescue routes and reducing unnecessary movement. Finally, workload assessment using the NASA-TLX scale revealed a significant decrease in perceived workload (t(df)=11.291, p=0.000, Cohen's d=2.525) (Figure 4), demonstrating the AR system's positive impact on reducing mental and physical strain. The reduced workload may result in decreased fatigue and improved long-term performance, benefiting the overall well-being of rescue personnel. These findings highlight the practical benefits of AR technology in increasing the efficiency, accuracy, and overall experience of rescue operations.

Decision-Making Process

Each participant was obliged to choose the same decision point for discussion in two experimental conditions: traditional map and AR-based search and rescue. Twelve participants decided to discuss the "marking buildings" decision-making process, while the remaining eight participants chose route planning as the major decision point for discussion. Uncertainty: 100% of the participants reported that they felt uncertain during the decision-making process owing to inadequate knowledge, time constraints, and the potential dangers linked with the repercussions of their choices. For example, one participant stated, "I continually checked if there were any stranded folks outside the house. I could view virtually the entire home via all of the windows, but I didn't notice any imprisoned people. I wasn't sure if I should identify this house as requiring additional searches or if there were any captive people. If I designated it red when there were no trapped persons, it would be a waste of our scarce medical resources. Another participant stated, "At the intersection, I needed to choose a bridge based on the map, but I had no idea which bridge was passable, and time was short. I needed to make the appropriate decision here so that I could search the entire impacted village more swiftly." This uncertainty becomes an obstacle to their decision-making, making it difficult for decision-makers to determine whether the outcome is "correct" or "helpful".

Decision Differences Between Two Rescue Methods: During the decision-making process for designating buildings, eight participants reported challenges with traditional map marking. For example, one participant commented, "After opening the map, I had to shift it left and right several times to locate the location of the building to be marked based on the coordinates and adjacent plants and structures. This put pressure on me." In the AR experiment, all 12 participants stated that after evaluating the building's damage condition, they just needed to select whether to designate the building as requiring further search, without having to constantly check the map, resulting in no decision-making issues.

Five of the participants who made route planning decisions reported feeling less pressure to make decisions throughout the AR experiment. Three participants reported that traditional map searches required more time to validate their position on the map and plan their next path. For example, one participant stated, "In the typical map experiment, I had to open the map numerous times and repeatedly check my location based on coordinates and the environment, as well as seek for adjacent structures that required searching. This was a huge waste of my time and energy. I couldn't see the house immediately when I was looking for the last structure on the other side of the river because trees were in the way. As a result, I checked the map five or six times during the search, wasting one minute. Ultimately, due to the ten-minute time limit, I failed to finish marking the building."

AR technology dramatically improves decision-making by delivering real-time information overlays and visual aids. When participants employ AR technology, they can make faster and more accurate decisions, saving time and cognitive strain from inadequate information and map scanning. This not only improves rescue efficiency but also increases decision accuracy and consistency.

Discussion

The findings of this study show that augmented reality (AR) technology offers substantial advantages over traditional rescue modes (TR) in earthquake rescue. First, AR technology shines at greatly lowering the workload of rescue personnel. According to the NASA-TLX questionnaire evaluation, participants using the AR system reported significantly lower workload across multiple dimensions than those using the traditional system, with Cohen's d reaching 2.525, indicating that AR technology has a significant impact on reducing cognitive and physical stress. This discovery is consistent with the research results of Alessa et al. (2023), which imply that AR technology can effectively reduce users' cognitive burden through real-time information overlay. Second, AR technology has considerably enhanced rescue efficiency, as indicated by shorter completion times and overall walking distance. Participants finished the task in an average of 201 seconds less while utilising the AR system, with a Cohen's d of 1.948, indicating a significant effect. Furthermore, the overall walking distance was reduced by roughly 65.36 meters, with an effect size of 0.719, indicating a modest impact. These results suggest that AR technology can enhance the overall efficiency of rescue operations by optimising rescue routes and reducing unnecessary movement, similar to the findings of Brizzi et al. (2018), who discovered that AR technology can significantly improve the speed and accuracy of task execution. In terms of decision-making, AR technology dramatically minimises the uncertainty that rescue professionals confront by delivering real-time spatial video feedback and information overlay, increasing decision-making speed and accuracy. Participants reported that AR technology enabled them to determine whether there were trapped individuals within buildings more rapidly, saving time and effort spent continually checking maps. This is congruent with the research conducted by Mitaritonna and Abásolo (2015), which demonstrates that AR technology can boost users' situational awareness and decision-making abilities, hence improving the overall task execution efficacy.

In addition, AR technology has improved the frequency and duration of map use. When participants utilise the AR system, they see fewer maps and spend less time looking at them. This not only improves the efficiency of information gathering, but it also reduces cognitive strain, allowing rescue personnel to concentrate on the actual rescue duties. However, in terms of accurately labelling the number of damaged structures, the AR system had a moderate influence (Cohen's d = -0.469), but its significance was only marginal. This finding requires additional research verification.

In conclusion, this study shows that AR technology has tremendous potential for use in earthquake rescue. It not only increases rescue efficiency and reduces workload, but it also improves decision-making. However, the study has some drawbacks, including a limited sample size and simulation limits in the experimental context. Future studies should increase the sample size and test AR technology in more realistic rescue scenarios to further validate its usefulness. Furthermore, the use of AR technology in various types of rescue activities and environmental elements should be investigated in order to completely grasp its benefits and limitations.

Impact of decision-making process

AR technology improves the decision-making process of rescue personnel by increasing situational awareness and information sharing. Participants noted that the AR system's real-time spatial video feedback enabled them to get a more thorough picture of the on-site scenario, eliminating uncertainty caused by limited information and allowing them to make more timely and correct decisions. This finding is consistent with study undertaken by Martins et al. (2022), which demonstrates that AR technology can boost users' decision-making efficiency and accuracy by giving instant environmental information.

Significant reduction in workload

The NASA-TLX questionnaire assessment revealed that AR devices greatly lower the perceived workload of rescue personnel. Specifically, after using the AR system, participants reported a considerable reduction in both cognitive and physical workload, with Cohen's d reaching 2.525. This finding suggests that AR technology has a high practical application value in relieving the work pressure and weariness of rescue personnel. Similar research has proven that AR technology can greatly reduce users' burden by delivering straightforward information display and lowering operational complexity (Alessa et al., 2023).

Significant improvement of rescue efficiency

The considerable reduction in completion time and total walking distance suggests that augmented reality technology can significantly improve the efficiency of rescue operations. Participants who utilised the AR system completed tasks faster and travelled shorter distances within the virtual world, demonstrating the efficiency of AR technology in optimising rescue routes and avoiding wasteful movement. This is consistent with the findings of Liu et al. (2023), which imply that AR technology can considerably enhance job execution efficiency through real-time navigation and route optimisation.

Optimisation of decision process

AR technology excels at optimising decision-making processes. When employing an AR system, participants may make judgements faster and more correctly, saving time and cognitive strain from frequent map checks.

Optimisation of map use

AR technology has dramatically reduced the frequency of map viewing and usage time, showing that rescue professionals can receive and use critical information more efficiently. This not only improves rescue efficiency but also reduces cognitive strain, allowing rescue professionals to concentrate more on the actual rescue duties. According to Qiu et al. (2024), AR maps are more effective than 2D maps for gaining spatial knowledge. AR displays are essential for helping people learn about sights and routes. The upgraded map broadens the scope of spatial information beyond the centre and highlights landmark sites to provide comprehensive information about the environment; hence, the improved map can aid in the acquisition of spatial knowledge.

Properly marked edges are significantly raised

Although the AR system had only a moderate influence on accurately labelling the number of damaged buildings (Cohen's d = -0.469), and its significance was just marginal (p = 0.050), this result suggests that AR technology has the potential to improve labelling accuracy. One possible explanation is that real-time information and visual support offered by AR technology assist rescue responders in more correctly identifying and labelling damaged buildings. However, because the p-value is merely on the edge of significance, future research should increase the sample size to confirm this effect.

Comprehensive discussion

Overall, this study shows that the use of AR technology in earthquake rescue offers considerable benefits. It improves rescue efficiency, decreases workload, and optimises decision-making. These findings are consistent with previous research, supporting the promise of AR technology in complex and dynamic contexts. However, the study has certain drawbacks, including a limited sample size and simulation limits in the experimental setting. Future studies should increase the sample size and test AR technology in more realistic rescue scenarios to further validate its usefulness. Furthermore, the use of AR technology in various types of rescue activities and under varied environmental conditions should be investigated in order to fully grasp its benefits and limitations.

Practical significance

The study's findings suggest that AR technology has tremendous potential for use in earthquake rescue. AR technology can significantly improve rescue efficiency by reducing task completion time and optimising rescue routes, allowing for the rapid rescue of more trapped individuals in emergency scenarios. Furthermore, the reduced workload reduces weariness and tension among rescue professionals, allowing them to perform more consistently while still maintaining their mental health. AR technology improves rescue personnel's decision-making ability and accuracy by overlaying real-time information and optimising decision-making points, eliminating erroneous decisions due to incomplete information. Simultaneously, eliminating unnecessary map reading and walking distance allows for more effective allocation of limited medical and rescue resources, thus boosting total rescue effectiveness.

Research limitations and future research directions

Although this study reveals the considerable benefits of AR technology in earthquake rescue, it also highlights certain limitations and suggests areas for future research. First and foremost, this study included a sample size of 20 participants. Although the results were significant, a bigger sample size study might help to validate and extend these findings. Second, this investigation was carried out in a virtual setting. In the future, AR technology should be tested in real or more realistic rescue scenarios to determine its true performance in complicated and dynamic surroundings. Furthermore, future research might look into the effects of AR technology in different sorts of rescue tasks and under varied environmental conditions to better understand its benefits and drawbacks. Future research should focus on the effectiveness of AR technology in long-term rescue missions, including its effects on the continuing work abilities and mental health of rescue professionals. Finally, investigating how to improve the functionality and user interface of AR technology to better suit the needs of rescue personnel in various scenarios will contribute to AR technology's practicality and application effectiveness.

Conclusion

This research confirms the potential of augmented reality (AR) to improve the efficacy of rescue operations, reduce workload, and improve decision-making. Compared to traditional approaches, AR-based tools allow participants to perform rescue missions faster while reducing the psychological (workload) burden. However, limitations of this study include the use of non-professional volunteers and the experiment's short duration, which may not accurately reflect the intricacy of real-world rescue attempts. Future study should include professional rescue teams, more complicated scenarios, and longer experimental periods in order to properly evaluate AR's usefulness and develop AR-based rescue tactics for practical use.

References

Alessa, F. M., Alhaag, M. H., Al-harkan, I. M., Ramadan, M. Z., & Alqahtani, F. M. (2023). A Neurophysiological Evaluation of Cognitive Load during Augmented Reality Interactions in Various Industrial Maintenance and Assembly Tasks [Number: 18 Publisher: Multidisciplinary Digital Publishing Institute]. Sensors, 23(18), 7698. https://doi.org/10.3390/s23187698

Brizzi, F., Peppoloni, L., Graziano, A., Stefano, E. D., Avizzano, C. A., & Ruffaldi, E. (2018). Effects of Augmented Reality on the Performance of Teleoperated Industrial Assembly Tasks in a Robotic Embodiment [Conference Name: IEEE Transactions on Human-Machine Systems]. IEEE Transactions on Human-Machine Systems, 48(2), 197–206. https://doi.org/10.1109/THMS.2017.2782490

Chalimas, T., & Mania, K. (2023). Cross-Device Augmented Reality for Fire and Rescue Operations based on Thermal Imaging and Live Tracking. Proceedings of the 1st Joint Workshop on Cross Reality (JWCR23) at ISMAR.

Demirkan, D. C., & Duzgun, S. (2020). An Evaluation of AR-Assisted Navigation for Search and Rescue in Underground Spaces. 2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), 1–2. https://doi.org/10.1109/ISMAR-Adjunct51615.2020.00017

Hu, D., Chen, L., Du, J., Cai, J., & Li, S. (2022). Seeing through Disaster Rubble in 3D with Ground-Penetrating Radar and Interactive Augmented Reality for Urban Search and Rescue [Publisher: American Society of Civil Engineers]. Journal of Computing in Civil Engineering, 36(5), 04022021. https://doi.org/10.1061/(ASCE)CP.1943-5487.0001038

Liu, M., Tang, L., & Zhou, J. (2023). The Efficiency and User Experience of AR Walking Navigation Tools for Older Adults. Augmented, Virtual and Mixed Reality Simulation. https://doi.org/10.54941/ahfe1004437

Macintyre, A., Barbera, J., & Smith, E. R. (2006). Surviving collapsed structure entrapment after earthquakes: A "time-to-rescue" analysis. Prehospital and Disaster Medicine, 21, 4–17. https://doi.org/10.1017/S1049023X00003253

Martins, N. C., Marques, B., Alves, J., Araújo, T., Dias, P., & Santos, B. S. (2022). Augmented reality situated visualization in decision-making. Multimedia Tools and Applications, 81(11), 14749–14772. https://doi.org/10.1007/s11042-021-10971-4

Mitaritonna, A., & Abásolo, M. J. (2015). Improving situational awareness in military operations using augmented reality [Accepted: 2018-04-11T12:36:42Z ISSN: 2464-4617]. Retrieved January 3, 2025, from http://dspace5.zcu.cz/handle/11025/29567

Qiu, X., Yang, Z., Yang, J., Wang, Q., & Wang, D. (2024). Impact of AR Navigation Display Methods on Wayfinding Performance and Spatial Knowledge Acquisition [Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10447318.2023.2169524]. International Journal of Human–Computer Interaction, 40(10), 2676–2696. https://doi.org/10.1080/10447318.2023.2169524

Wang, M., Ji, H., Jia, M., Sun, Z., Gu, J., & Ren, H. (2023). Method and application of information sharing throughout the emergency rescue process based on 5G and AR wearable devices [Publisher: Nature Publishing Group]. Scientific Reports, 13(1), 6353. https://doi.org/10.1038/s41598-023-33610-4

Xu, F., Zhou, T., Nguyen, T., & Du, J. (2024). Augmented reality in team-based search and rescue: Exploring spatial perspectives for enhanced navigation and collaboration [Publisher: Elsevier]. Safety Science, 176, 106556.

Function Allocation for Responsible Artificial Intelligence: How do we allocate trust and responsibility?

Patrick Waterson[1], Chris Baber[2], Edmund Hunt[3], Sanja Milivojevic[4], Sally Maynard[1] & Mirco Musolesi[5]

[1]Human Factors and Complex Systems Group, Loughborough University, [2]University of Birmingham, [3]University of Bristol, [4]Bristol Digital Futures Institute, University of Bristol, [5]University College London

SUMMARY

We consider how guidelines for Responsible Artificial Intelligence (RAI) need to be adapted to address the challenges of Function Allocation (FA) in human-agent teams. We offer an approach that takes a system description, using CWA, to identify where responsibility for consequences of actions might lie across the system. We propose that, in addition to allocation of functions, analysis of the system needs to identify decision points (where agents have a choice of action to perform) and responsibility points (where agents identify the consequences of their decisions). We illustrate this with example experiments. We put forward a set of open challenges and questions facing researchers in the areas of RAI and FA. We point to the need for greater emphasis on the issue of responsibility, trust and accountability in new forms of automation. We also provide pointers for the future and how these might be addressed in the coming years.

KEYWORDS

Function allocation (FA); Human-Robot teams; Responsible Artificial Intelligence (RAI); sociotechnical systems.

Introduction

Prominent computer scientists have questioned the ethical, moral and legal implications of the new technologies built of 'foundation Artificial Intelligence' (e.g., Brown, 2023). These implications often lead to calls for so-called 'responsible AI (Artificial Intelligence).' A key aspect of this, for Human Factors, is how responsibility relates to Function Allocation. That is, in a system involving AI agents and humans, where does responsibility for decisions and actions sit? Is it appropriate to assume that only humans can hold responsibility for decisions or should the human responsibility be for the consequences of these decisions? Does this responsibility for consequences of decisions apply even in circumstances where the behaviour of the AI is not transparent? Function allocation (FA) refers to strategies for distributing system functions and tasks across people and technology. Traditionally, as discussed below, FA has concentrated on comparisons between humans and machines in terms of their strengths and weaknesses. With the advent of more advanced technologies such as artificial intelligence (AI) and robotics, it can be more difficult to identify clear boundaries between capabilities of humans and agents (particularly in tasks involving decision making) and there is a need to move beyond conceptions of FA as discrete allocations and to move towards new forms of interdependencies that will create shared, collective entities involving humans and machine. Many FA methods pay limited attention to joint operation of people to perform tasks, and very few FA methods address Machine-Machine (M-M) allocations, particularly in terms of how much autonomy should be provided to technologies such as robots and agents.

Traditional approaches to function allocation

The topic of function allocation can trace a lineage back to some of the earliest days of HFE and the seminal work of Paul Fitts (1951) on HABA-MABA lists (Humans-are-better-at-Machines-are-better-at). At the time Fitts and his colleagues were developing ways of examining automation, well before the computer and internet revolutions of the 1970s through to today. In simple terms, 'automation' will operate according to clearly specified sets of instructions; systems with semi-autonomy will follow instructions but can adapt their behaviour according to situation; autonomous systems have the potential to define their own instruction sets and specify their goals. We believe that much of the FA literature has focused on automation, with some effort at considering semi-autonomous systems, and there remains a gap when it comes to FA for autonomous systems that interact with humans. Following Fitts' initial proposals, approaches to FA stressed the need to encompass wider sociotechnical aspects of work design and a wider range of concerns (Waterson *et al.*, 2002, figure 1).

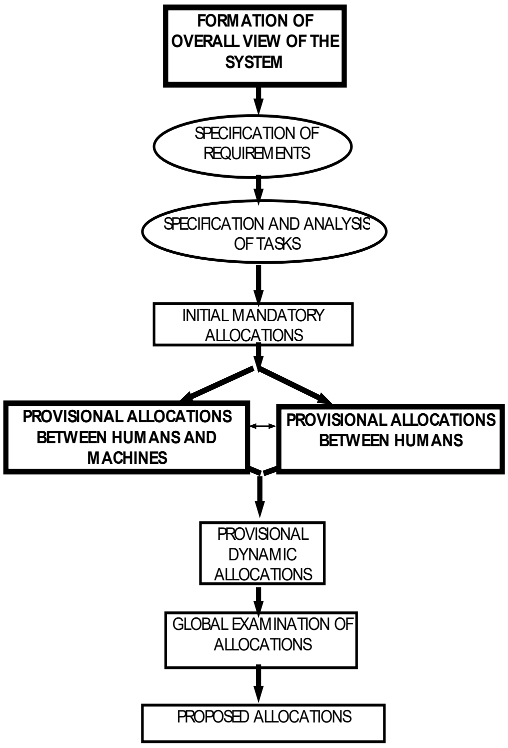


TABLE 3: Decision Criteria for Use with the Method

Category of Criteria	Criteria
Goal issues	Goals of the system, goals of the organization, other goal issues
Organizational issues	Existing/required systems and procedures, existing/required organizational norms issues and practices, organizational requirements, organization structures, accountability, other organization issues
Cultural/environmental issues	Cultural issues, legal requirements, health and safety, data security, national/international political considerations, other cultural/environmental issues
Resource issues	Technology, people, money, other investments (e.g., time), knowledge, skills, education, other resource issues
People issues	Level, type, and number; knowledge, skills, and education; trust; training; social acceptability; other people issues
Task issues	Speed of response, accuracy, frequency, physical demands, cognitive demands, emotional demands, reliability, efficiency, flexibility, operational criticality (likelihood and implications), safety criticality, uncertainty of occurrence, uncertainty of the type of problem, uncertainty of the correct response, variability in performance, task interdependencies, redundancy, other task issues
Job design/work organization issues	Control (e.g., autonomy, responsibility, variance handling, decision making), accountability, variety, skill use and development, workload, clarity of goals and requirements, feedback on performance, communication and social contact, health and safety (mental and physical), flexibility, information exchange and knowledge sharing, situation awareness, understanding of the system, mutual adjustment/team collaboration, role ambiguity and overload, motivation, satisfaction, stress, performance, other job design/work organization issues
Technology issues	Feasibility and cost of automation; maintainability; reliability; level of performance; trust in technology; level, type, and amount of technology; other technology issues

Figure 1: Flowchart and decision -criteria for function allocation (Waterson et al., 2002)

Existing approaches to FA might be limited by their assumptions about the capabilities of technology, but the emphasis of FA occurring within a Sociotechnical System is important to future developments of the concept. However, even though existing FA approaches address the distribution of work across a wider system than solely one human and one machine, they tend to underplay the role of trust, ethical, legal and wider societal issues. In broad terms, these considerations fall under the heading of 'responsible innovation' and raises questions of how-to responsibility and where it sits in a sociotechnical system. With the advent of sophisticated technology that allows humans to operate in teams with these technologies, such as Human-Robot

Teams (HRTs) and Human-Agents Teams (HATs), there is an increasing potential for semi-autonomous and fully autonomous working arrangements.

Responsible AI and FA – shifting the focus

In this paper we focus on interdependencies in human-automation systems and how these might be framed by a concept of responsible FA. Interdependencies play a large part in recent approaches to FA. However, we propose that current approaches to FA focus on the capability of actors and (sometimes) predictability of outcomes. FA does not, generally, address situations when Actors are pursuing competing, contradictory, erroneous goals. Likewise, robots, and semi/fully-autonomous systems are capable of defining their own intent and this intent might differ from that of their human team-mates. In this case the question remains how can FA address the challenges this raises?

Levels of Responsibility

Levels of Automation indicate interdependencies between human and automation, depending on the capability of the automation. From levels of automation, a stage model of responsibility could be proposed, e.g., an agent can perform the action or make a decision within limits defined by its design or user; an agent can suggest action or decision, but the human is responsible for performing this, an agent can suggest decisions or action etc., and human can veto; an agent cannot do action etc. In addition to defining levels of automation in terms of capability, Wickens et al. (2010) added types of cognitive task, e.g., attention (i.e., acquiring information), information integration (i.e., combining information from different sources), decision (i.e., choosing an action or interpretation to perform on the basis of the information), execution (i.e., performing the action or interpretation). This, we feel, allows us to highlight the ways in which FA can apply across human-agent teams, and also suggests likely points at which responsibility can be delegated from human to agent. For example, assume that a (firefighter) Incident Commander is working in a team with uninhabited aerial vehicles (UAVs), and has set these the objective of reconnoitring a large, open plan building such as a factory or warehouse. Assume further that the building is filled with smoke and the objectives involve search for the source of a fire and to find combustible materials. Setting these objectives, the Incident Commander might assume that the UAVs have sensor and navigation capabilities that are sufficient to enable the objectives to be met. Rather than the UAVs continually checking with the Incident Commander as to whether their interpretations are correct, we believe it more likely that they would provide a running update of situation awareness to enable the Incident Commander to develop a response to the fire and the risks identified. The challenge arising from this, is as follows: if the UAVs fail to detect something in the environment that has an impact on the plan, and the impact results in catastrophic consequences, does the responsibility for this lie with the Incident Commander? And, if the responsibility does not lie with the Incident Commander, in what ways are the UAVs culpable (or should responsibility lie with their designers, manufacturers, distributors, maintainers, etc.)?

Approach

Starting from a revision of the FA framework of Waterson et al. (2022) we consider how we might incorporate decision criteria that address the issue of responsibility, and expansion of the sociotechnical system beyond human-machine dyads. This provides an impetus for considering not only decision points in the definition of a task (e.g., where the task involves the choice between actions in response to the situation), but also 'responsibility points' (e.g., where the outcome of the decision becomes identifiable and can be evaluated in terms of acceptability). This contrast between decision and responsibility points in the timeline of a mission could be considered a priori in mission planning. For example, in the Incident Commander example above, decision points could relate to navigation, sensing, interpretation. It could be feasible to assume that the semi-autonomous UAVs would be capable of acting appropriately at these decision points. However, the responsibility points might involve the decision to change firefighting tactic which might alter the risk to personnel. In this respect, the mission plan would define which decision outcomes need to be evaluated in terms of acceptable risk.

The notion is that there are responsibility points (that operate by analogy with decision points). At these responsibility points, there are additional sociotechnical systems considerations. First, which part of the sociotechnical system defines 'acceptable risk'? Second, which part of the sociotechnical system predicts the likelihood of outcomes from specific decisions? Third, which part of the sociotechnical system evaluates the decision outcomes?

The definition of 'acceptable risk' could be made in advance of specific operations, probably through Standard Operating Procedures, and that responsibility for this definition lies in the organisations command hierarchy. That is, if an agent demonstrably follows these SOPs and demonstrably responds within the agreed limits of acceptable risk, then responsibility ought to lie with the organisation rather than the individual. We recognise that such an argument raises legal and ethical concerns (particularly in terms of the assumption that the SOP would be appropriate to a given situation). However, starting from this assumption means that we can define responsibility points in a mission. It also means that computer agents could be tasked with decisions and actions that operate within the SOPs. While we are not arguing that the agents are responsible, it does imply that they can potentially be tasked with decisions and actions that have outcomes which could affect 'acceptable risk'. However, this is not to claim that the agents are engaged in moral or ethical decision making. Rather, it is to assume that the SOPs contain within them the moral and ethical imperatives that will constrain decisions and actions. Having outlined an argument as to how computer agents could make such decisions, we now want to argue against this. The reason for urging caution on permitting agents to make decisions that affect responsibility points it that one would need to guarantee that the situation is entirely defined in a way that fits the SOPs. In any complex incident, it is likely that the situation will unfold in ways that are unanticipated. Consequently, rather than allowing the agent the opportunity to make such decisions, these would always need to be referred to a human in the team who held sufficient authority and would be prepared to take responsibility for the outcomes. This, in turn, requires sufficient trust in all actors (human and computer) engaged in decisions that affect these responsibility points.

The discussion so far has implications for what is defined as a 'function' within FA (i.e., where the agent, human or machine, is responsible for the outcome, not just who does something, of an action of decision), as well as the wider issue of trust in automation. We define trust in the manner of Lewis and Marsh (2022) who construe it in terms of the following components:

- 'Trust' involves a 'model' of teammates held by members of a team, defined by:
    - Capability: is an agent (or human) is able to perform a given function at a given point in time (in its own or its team-mates opinion) and is it available?
    - Predictability: what is the probability of success of completing the function, and will it do this is way that team-mates expect?
    - Integrity: will the team-mate pursue individual or team goals, and will performance by within moral, legal, and ethical constraints?

In our ongoing work, we argue that 'trust' is dynamic and will need to be sufficient to support collaboration with the team, i.e., trust is satisficed (Baber et al., 2023; Hunt et al., 2024). This means that the members of a team ought to sufficiently confident that their team mates are acting in the interests of the wider sociotechnical system (as far as possible) and are able to identify which team members are involved in specific decision points and responsibility points.

Initial Studies

As a simple 'proof of concept' of the notion of responsible FA, we built an agent-based model using NetLogo model (Baber et al., 2023). In this model, 3 Agents explore a Maze. Each Agent has specific Functions that include navigating the maze (we define this using simple functions, such as left-hand-on-wall where the robot will follow a wall unless it is able to turn left, and if it reaches the end of a line it turns left until it can move forward). In addition to navigation, each agent has a specific function. In this case the yellow Agent (Figure 2) is 'trapped' by a red square and needs to be 'rescued'. For the blue agent, the red square is a token to be collected. Collecting the token would result in the coincidental 'release' of a trapped yellow agent. The green agent, needs to seek trapped agents and release them from the red square and 'rescuing' other agents is a priority.

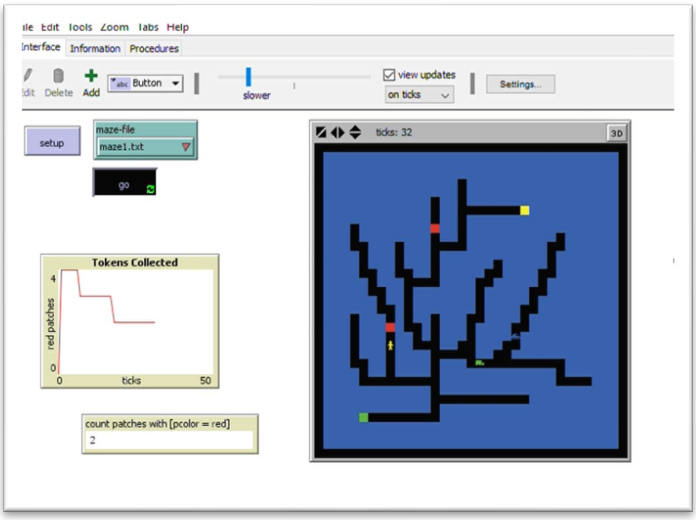


Figure 2: NetLogo Model – 3 Agents exploring a Maze (Baber et al., 2023)

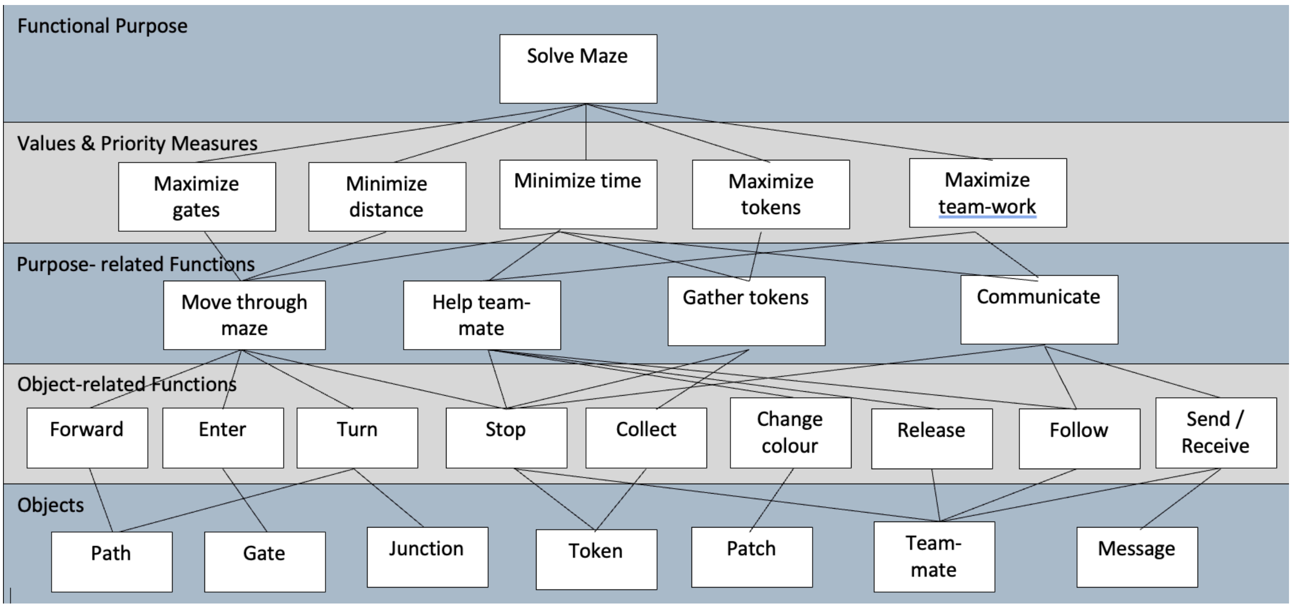


Figure 3: Work Domain Analysis diagram sketching out a description of the constraints that govern the purpose of the agents and the function of the systems as a whole.

From Figure 3, we can define several decision points that can be deduced from the Object-related Functions. For example, if it is not possible to move forward, then turn then; if there is a gate on the left, then enter it; if there is a junction, then take the left turn. Further, we can define several responsibility points from the Purpose-related Functions. For example, if a team-mate is trapped then stop your current task and release them; if tokens are available, then collect them; if you are trapped, then communicate with your team-mates. In terms of responsibility, we propose that the consequences of decisions would be internal to the team and do not affect a wider sociotechnical system. This means that if the choice taken at a responsibility point is not for the good of the team (i.e., continuing to collect tokens to boost your own score, rather than stopping to help a team-mates) then the outcome only affects the performance of the team.

One of the interesting outcomes from the simulation was that if an agent changes behaviour (e.g., becomes less attentive to details, fails to identify risks, prioritises non-altruistic goals, etc.), the overall outcome (a failed mission) is a consequence of a shared failure of the team, and responsibility is also shared. The action of team-mates serves to constrain the choices available to individual agents. In this context, overall responsibility exists ‘between’ team-mates.

Experiments in Human-Robot Teams

The second part of our paper concern a set of experiments which were designed to manipulate trust in human-robot teams (HRTs). We analysed data from HRT experiments focused on trust dynamics in teams of one human and two robots (Figure 4 shows an example of a participant working with the floor-based robots), where trust was manipulated by robots becoming temporarily unresponsive. In this case, the mission was to collect tokens (jointly or independently with the robots) and lack of response from the robot compromised the ability to jointly collect tokens. This compromised ability, while it was the result of a deliberately induced technical failing, was often perceived by the human participants in terms of the robots making a choice. For example, the robot was perceived to not want to help the human or was perceived as being too busy performing its own tasks to offer help. From this, the robot was ascribed some degree of agency that allowed it to make a choice as to whether or not to help. The choice could be considered as a decision point, but could also be interpreted in terms of a responsibility point. That is, if the mission was for a team to gain a certain number of points and (due to lack of responsiveness) the team was not able to meet the required number, the robots could be perceived to be responsible for the team’s poor performance.

Figure 4: Experimental set up – participant and robots (based on Hunt et al., 2024).

Discussion

Our research so far shows some efforts to shift away from traditional FA and raises many questions about agency and responsibility in human-machine and robotic systems in general. There are currently no established HFE-oriented criteria or structured questions regarding how we might allocate 'responsibility' between humans, machines, and numerous forms of new technology (e.g., robots, drones). Likewise, the interdependencies which might exist in human-machine interaction are only starting to emerge (e.g., robotic systems which have the ability to 'trust' their human operators). These sorts of concerns raise a number of thorny, philosophical questions (e.g., how to allocate 'blame' and 'accountability' when something goes wrong when humans and machines jointly carry out a task). Considering FA in terms of decision points and responsibility points (in addition to considering functions) and then considering how humans and agents in a team are entrusted to respond to such points, provides a first step in incorporating responsibility into FA.

Acknowledgements

The research reported in this paper is supported by grant EP/X028569/1 'Satisficing Trust in Human Robot Teams' from the UK Engineering and Physical Sciences Research Council (EPSRC). This project runs from 2023 to 2026 and involves the Universities of Birmingham, Bristol, Loughborough and UCL.

References

Baber, C., Waterson, P.E, Milivojevic, S., Maynard, S., Hunt, E.R. and Yusuf, S. 2023. Incorporating a 'ladder of trust' into dynamic Allocation of Function in Human-Autonomous Agent Collectives. In the 14th Organizational Design and Management Conference (ODAM). Bordeaux, France

Brown, S. (2023). Why neural net pioneer Geoffrey Hinton is sounding the alarm on AI. Sloan Management Review, May 23, 2023.

Fitts, P. M. (1951). Human engineering for an effective air navigation and traffic control system. Washington, DC: National Research Council.

Hunt, E. *et al.* (2024). Co-Movement and Trust Development in Human-Robot Teams. https://doi.org/10.48550/arXiv.2409.20218.

Jian J., Bisantz A. & Drury C. (2000). Foundations for an empirically determined scale of trust in automated systems, International Journal of Cognitive Ergonomics, 4, 53–71.

Lee, J.D. and Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. International Journal of Human-Computer Studies, 40(1), 153–184.

Lewis, P.R. and Marsh, S. (2022). What is it like to trust a rock? A functionalist perspective on trust and trustworthiness in artificial intelligence, Cognitive Systems Research, 72, 33-49

Mayer, R.C., Davis, J.H. and Schoorman, F.D. (1995). An integrative model of organizational trust, The Academy of Management Review, 20, 709-734

Waterson, P.E., Older-Gray, M. and Clegg, C.W. (2002). A sociotechnical method for designing work systems. Human Factors, 44, 3, 376-391.

Wickens, C.D., Li, H., Santamaria, A., Sebok, A. and Sarter, N.B. (2010). September. Stages and levels of automation: An integrated meta-analysis. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 54, No. 4, pp. 389-393). Sage CA: Los Angeles, CA: Sage Publications.

Predicting Compliance Behaviour During a Flood Disaster Using the Talk-Through Method

Razan Y. Aldahlawi[1,2], Glyn Lawson[2] & Vahid Akbari[3]

[1]Industrial and Systems Engineering Department, Faculty of Engineering, PNU, Saudi Arabia, [2]Human Factors Research Group, Faculty of Engineering, University of Nottingham, UK, [3]Operations Management and Information Systems Department, University of Nottingham, UK

SUMMARY

This study was conducted to explore compliance behaviour during flood evacuation, employing the talk-through approach developed by Lawson (2011), for which participants are required to explain their anticipated actions in response to a hypothetical emergency scenario. The results showed that social cues, as described by the Protective Action Decision Model (PADM) (Lindell & Perry, 2012), were found to be the most effective factor in reducing evacuation delay time and enhancing compliance with evacuation orders. It was also found that in the absence of any routing instructions or specific cues, evacuees ranked familiar and shortest (by distance) routes as more preferable than those taken by neighbours. This study contributes to our understanding of evacuees' delay times and compliance with evacuation instructions, under the influence of different factors, including environmental cues, social cues, warning message contents, and various sources of the warning. It also demonstrated the value of the talk-through as a tool for studying human behaviour in floods without the risks or complications associated with data collection in real flood events.

KEYWORDS

Human factors, flood evacuation planning, compliance behaviour

Introduction

Compliance behaviour is a significant factor contributing to successful evacuation operations (Wang, et al., 2020). Noncompliance with instructions would increase casualties during large-scale movements, specifically in stressful and degraded environments (Daudé, et al., 2019). To help emergency planners and managers create efficient strategies, more empirical research is needed for a more thorough understanding of the factors that contribute to non-compliance (Saha & James, 2017). Additionally, most prior research has focused on how pedestrians respond during a building evacuation; less emphasis has been placed on how people behave while driving a vehicle during evacuation in a large-scale emergency (Bakhshian & Martinez-Pastor, 2023; Aldahlawi, et al., 2024). This work builds upon prior research by addressing evacuees' compliance with instructions during a hypothetical flood scenario. It presents a holistic overview of how people will likely comply in both the traffic network and within buildings, providing valuable insights into behaviour throughout all phases. Overall, this study was conducted to answer the question "What are the key factors influencing evacuees' compliance behaviour with evacuation departure time and routing instructions?", testing the following hypothesis:

H_1: Evacuation delay will vary depending on various influencing factors, including environmental cues, social cues, evacuation distances, levels of congestion, and sources of warning.

H_2: The likelihood of compliance with evacuation departure time will differ across the abovementioned conditions.

H_3: The likelihood of compliance with evacuation routing will differ across the conditions.

Method

Participants

The participants included 26 Saudi citizens (Mean age: 33; range: 23-44) living in areas prone to flooding. Of the total sample, 12 were female and 14 were male. The average household size was 6 members (range: 2-29). Participants reported a median awareness level of flood risks and flood resilience actions in local areas equal to 3 (1=very low, …, 5=very high). Responses show that most participants live in two-story buildings. People with severe health conditions, mental health issues, traumatic experiences of a flood, or lost a close friend or relative in a disaster were excluded from this study. The recruiting process included sending an advertisement via social media.

Materials

The talk-through method developed by Lawson (2011), for which participants are required to explain their anticipated actions in response to a hypothetical emergency scenario, was used to measure the influence of the PADM variables (i.e., environmental and social cues, information sources, and message contents) on compliance behaviour. A questionnaire was developed consisting of 14 Likert-type scale questions to measure participants' likelihood of complying with evacuation orders. Higher scores indicated a higher level of compliance. Rating scales were chosen as they are the most frequently used method in subjective assessment, allowing the participants to express their perceptions (Wilson & Sharples, 2015).

The first section of the talk-through questionnaire focused on demographic and socio-economic information. Demographic information such as age, gender, household size, relationship with the household members, and awareness of flood risks and flood resilience actions in the local area were collected, describing the characteristics of the study sample. Likewise, socioeconomic information involving residence type was acquired. After that, participants were given the time to draw their property layout, as Lawson described (2011). Afterwards, they were presented with a hypothetical flood evacuation scenario and asked to answer several questions. Figure 1 shows the hypothetical flood evacuation scenario.

Imagine that you and all your household members are currently at home. You receive the following text message from the government: “Immediate evacuation is needed due to upcoming flooding. Your home is in the highest-priority evacuation zone. Use your own vehicle and follow Route 1 immediately. Go to the assigned Shelter A for safety.” The image below is also sent by text message, although it will contain more detailed local area information regarding routes and destinations:

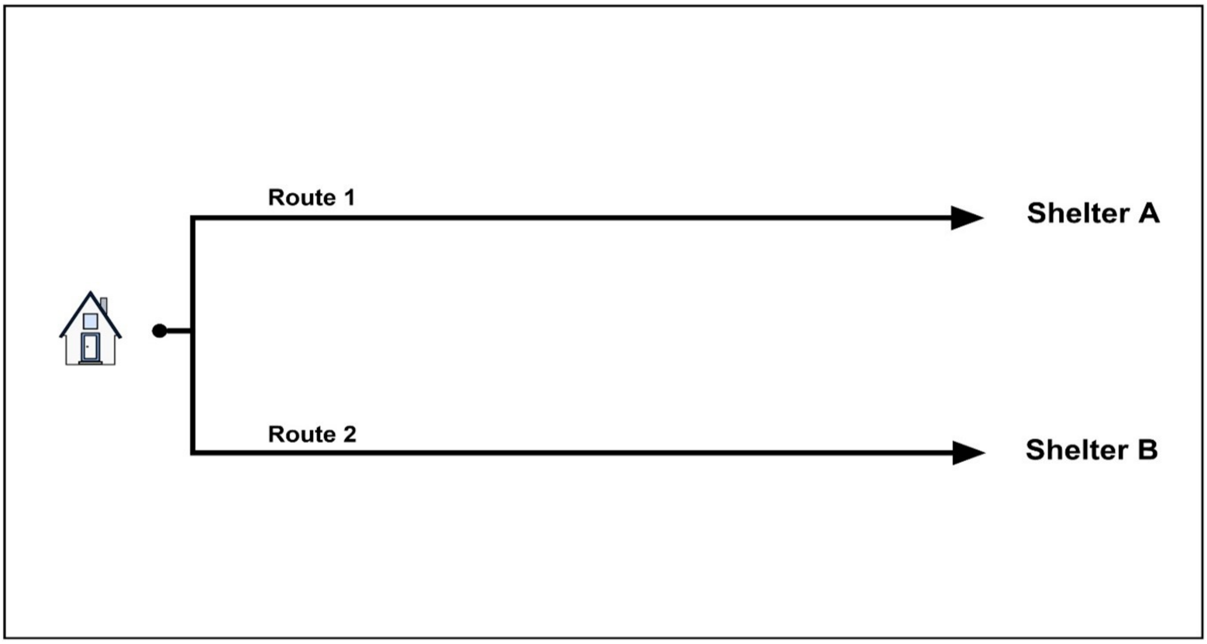


Figure 1: The hypothetical flood evacuation scenario employed in the talk-through questionnaire

The talk-through questions included what actions they would take upon receiving the warning until they reached the shelter, specifying the location of each action. Participants were also asked to predict their expected departure time following the evacuation warning (i.e., control condition for studying the evacuation delay). Then, participants were asked to rate the influence of the PADM variables on their compliance with evacuation departure time, as well as predict their anticipated departure time for each condition. Following that, participants were asked to rank their preferences regarding the evacuation route from six route characteristics, including most familiar, least congested, shortest by distance, on high ground, recommended by Google Maps, and taken by neighbours. Participants were also given the option of stating their preferred (i.e., distinct from the given list) route and ranking it. However, this data was excluded from the analysis, because only four participants reported a preferred route, which was an inadequate number of responses to be considered. At last, participants were asked to rate the influence of the PADM variables on their compliance with evacuation routing instructions.

Procedures

PADM is a theory that can be employed as the foundation for designing a conceptual model of human behaviour and decision-making during the time before evacuation in an emergency (Lindell & Perry, 2012). According to Lindell et al. (2019), PADM is a suitable framework for studying human behaviour during a flood evacuation. Following a within-subject design, this study explored how the different PADM-derived conditions affect evacuation delay time, compliance with evacuation departure time, and compliance with routing instructions. The PADM-derived conditions employed to study evacuation delay and compliance with departure time are as follows:

Condition 1: Noticing heavy rainfall accompanied by gradually rising, visible, flood level.

Condition 2: Seeing neighbours staying at home (noncompliance influence).

Condition 3: Seeing neighbours evacuate (compliance influence).

Condition 4: Receiving information on a shorter evacuation distance to the shelter (2.5 km).

Condition 5: Receiving information on a longer evacuation distance to the shelter (5 km).

Condition 6: Receiving information indicates heavy congestion with the warning message.

Condition 7: Receiving information indicates moderate congestion with the warning message.

Condition 8: Receiving the evacuation warning from an official government siren, instead of by text message.

Condition 9: Receiving the evacuation warning from social media instead of receiving it from official government sources.

Similarly, the PADM-derived conditions employed to study compliance with routing instructions are as follows:

Condition 1: Noticing heavy rainfall accompanied by gradually rising, visible, flood level.

Condition 2: Seeing other vehicles take the suggested route per the evacuation instructions (compliance influence).

Condition 3: Seeing other vehicles take a different route to the evacuation instructions (noncompliance influence).

Condition 4: Noticing congestion on the instructed route.

Condition 5: Availability of another evacuation route, distinct from the instructed one, which is more familiar.

Considering that the factors influencing driving behaviour, including route familiarity and traffic conditions (Wu, et al., 2012; Bian, et al., 2023), differ from those associated with pedestrian behaviour, PADM-derived conditions were modified as appropriate. The flood evacuation scenario included receiving an evacuation warning by a text message while at home, hence, multiple warning sources (i.e., social media or official siren), along with being informed about evacuation distances (i.e., shorter or longer distances) and traffic conditions (i.e., moderate or heavy congestion) presented in the warning message, were excluded from studying driver compliance.

The talk-through sessions were conducted online or in person, depending on the participant's preference and availability. A partial counterbalancing strategy was applied, producing two forms in which selected questions were reversed to minimise order effect. To elaborate further, in the "message content" conditions, half of the participants answered the "heavy congestion" condition first and, after that, answered the "moderate congestion" condition, whereas the other half did the opposite. Also, half of the participants answered the "shorter evacuation distance" condition first and, after that, answered the "longer evacuation distance" condition, whereas the other half did the opposite. However, other conditions were not counterbalanced which is considered a limitation to this study. Future studies should apply a more efficient counterbalancing strategy in a within-subject design to control bias. The participants were randomly assigned to one of the forms.

Results

As mentioned, a taxonomy based on the PADM framework was developed to analyse the results.

Evacuation Delay Behaviour

The repeated measures ANOVA showed a significant difference in evacuation delay time between the different conditions (F (1,13) =22.546, p <0.001). Paired samples t-tests showed substantial differences between the control condition, in which participants were told they received a warning by text, but no other influences were mentioned, and both social cues (compliance and noncompliance influence) conditions. The social–compliance influence condition showed a substantial decrease in delay time compared to the control condition (mean difference = 5.65, p = .003). In contrast, the social–noncompliance influence condition showed a substantial increase in delay time compared to the control condition (mean difference = 3.77, p < .001). Figure 2 illustrates the average evacuation delay time (i.e., in minutes) predicted by the participants across the PADM-derived conditions.

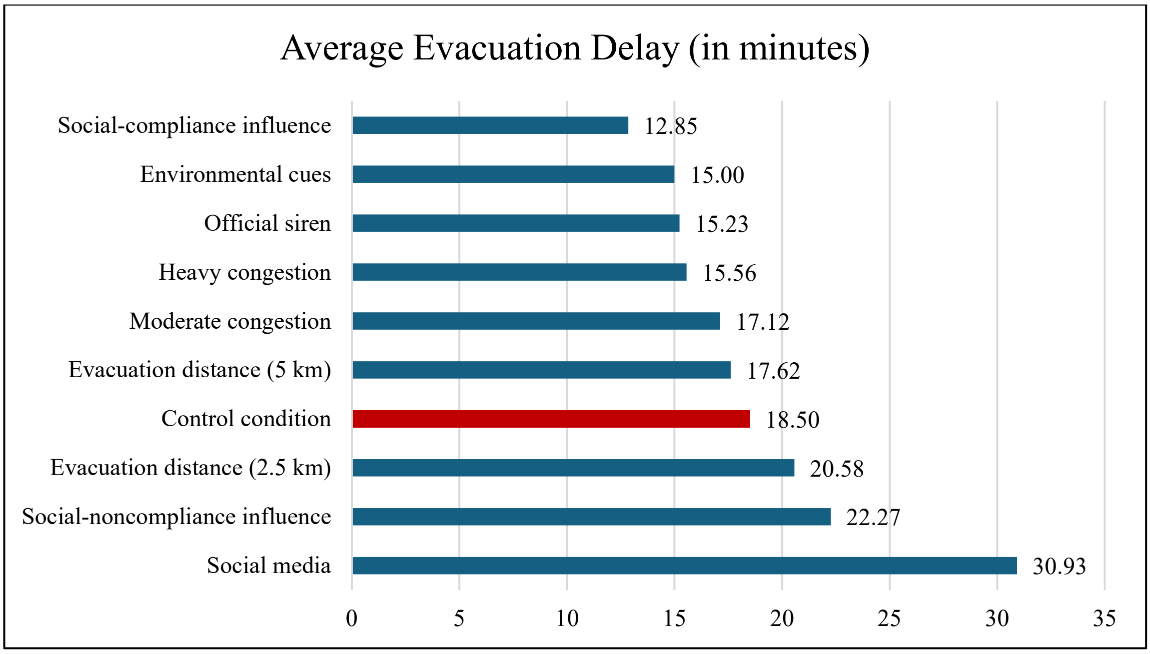


Figure 2: Comparison between the evacuation delay averages in the different conditions

As appeared in Figure 2, the average evacuation delay time was 18.50 minutes in the control condition, highlighted in red. The shortest delay time, reported in the "social-compliance influence" condition, was 12.85 minutes, while the longest anticipated delay time was 30.93 minutes when social media delivered evacuation warnings.

Compliance behaviour

Friedman's tests were employed to evaluate participants' ratings of their anticipated compliance with evacuation departure time and routing instructions across the different conditions. The results indicated significant differences between the conditions in both compliance with departure time ($\chi^2(8) = 87.218$, p<0.001) and routing instructions (χ^2 (4) = 48.877, p<0.001). The Wilcoxon Signed-Rank tests, applying Bonferroni correction, revealed significant differences across many condition comparisons.

For departure time compliance, receiving warnings from social media led to significantly lower compliance compared to social–noncompliance influence (p = 0.001), environmental cues, social–

compliance influence, evacuation distances, and congestion levels (all: $p < 0.001$). Additionally, social-noncompliance led to significantly lower compliance compared to being told about a longer distance to evacuate (5 km), environmental cues (both: $p=0.001$), social-compliance influence, and official siren (both: $p < 0.001$). Finally, being told about a shorter evacuation distance (2.5km) led to significantly lower compliance compared to social-compliance influence and environmental cues (both: $p < 0.001$). Figure 3 illustrates the median values for the rating scores across the PADM-derived conditions.

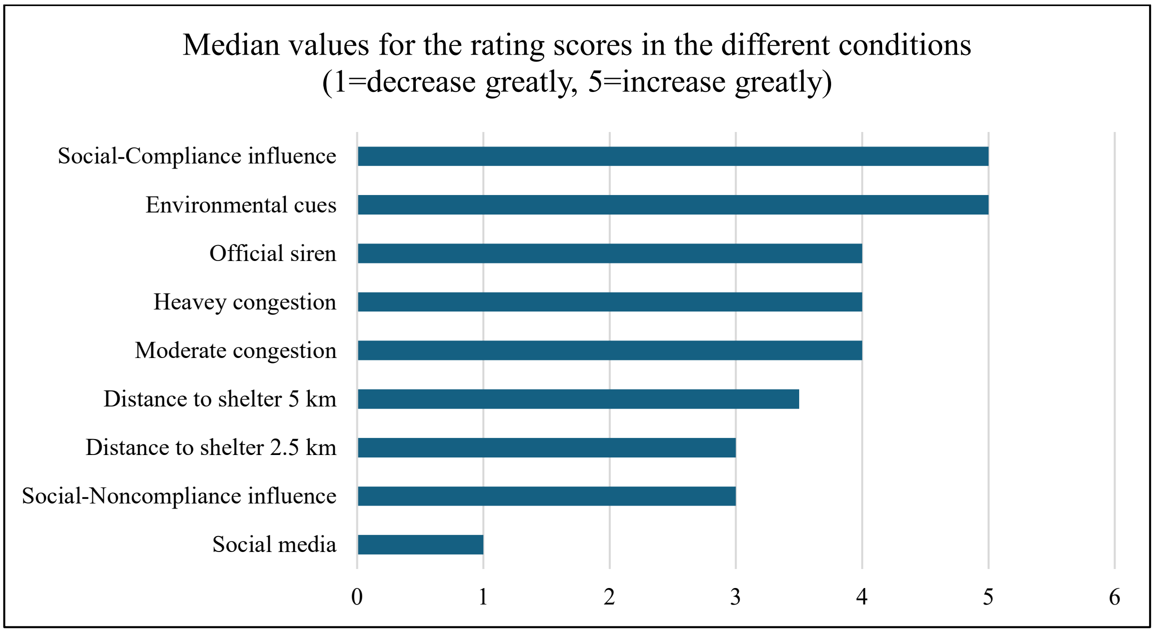


Figure 3: Median values for the rating scores measuring the influence of the PADM variables on compliance with evacuation departure time

For routing instructions compliance, social-compliance influence resulted in greater compliance compared to environmental cues ($p < 0.002$), social-noncompliance influence, congested routes, and familiar routes (all: $p < 0.001$). Furthermore, familiar routes led to a marked reduction in compliance relative to environmental cues ($p < 0.004$). Figure 4 illustrates the median values for the rating scores across the PADM-derived conditions.

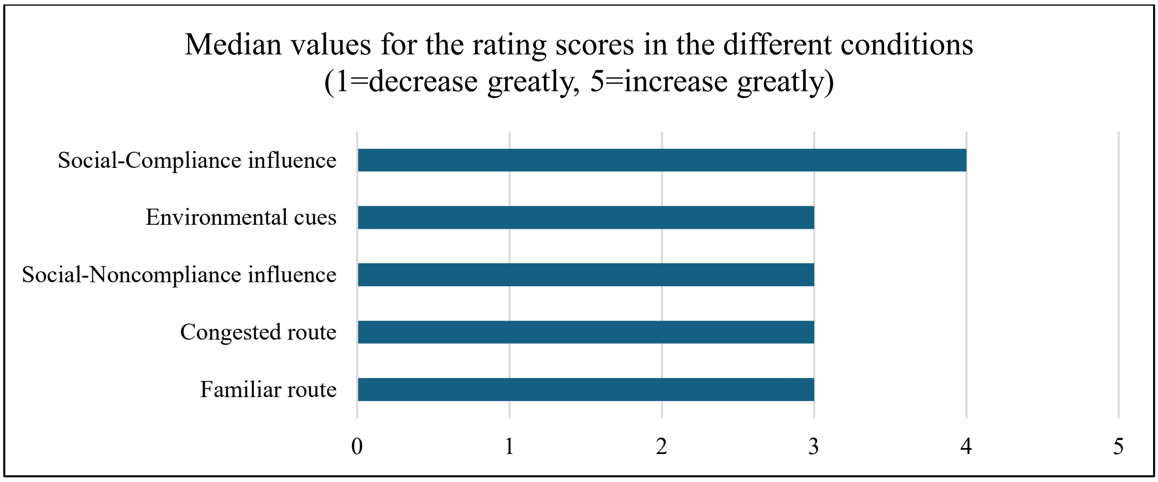


Figure 4: Median values for the rating scores measuring the influence of the PADM-derived variables on compliance with evacuation routing instructions

Evacuation routing preferences in floods

Participants were asked to rank their preferred evacuation route from a given list of route characteristics (familiar, least congested, shortest by distance, on high ground, Google Maps, and neighbours take). Ranking scores take values from 1 to 6, in which 1 reflects the most preferred evacuation route. A Friedman's test was employed to evaluate ranking values across the different evacuation routing characteristics. The results indicated a significant difference between the conditions ($\chi^2(5) = 19.125$, $p < 0.002$). The Wilcoxon Signed-Rank tests, with Bonferroni correction employed, showed that evacuees significantly preferred familiar routes and the shortest (by distance) routes over those taken by neighbours ($p = 0.002$ and $p = 0.003$, respectively). Figure 5 illustrates the median values of the rankings of evacuation routing preferences.

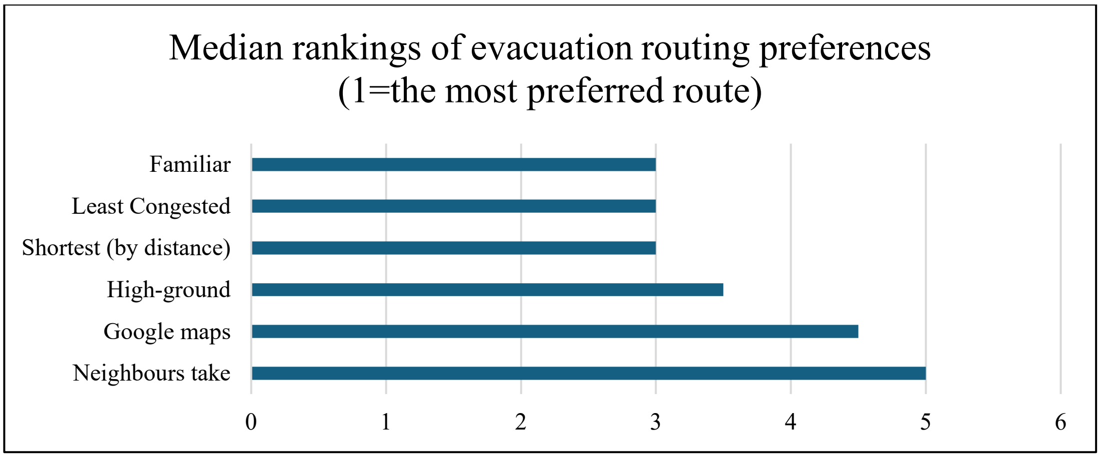


Figure 5: Median values of the rankings of evacuation routing preferences

Remarkably, data analysis suggests that social cues are highly influential on compliance behaviour and time to leave; however, when it comes to preference, evacuees would prefer to take familiar or shortest (by distance) routes. This result is consistent with previous findings, in which route choices are based on the routes located within a short distance or well-known routes (Charnkol et al., 2007; Sadri et al., 2014; Whitehead et al., 2001). Additionally, route familiarity and route choice influence driving compliance greatly, as evacuees tend to follow familiar roads and motorways in emergency evacuation (Chiu & Mirchandani, 2008).

Behavioural pattern in flood evacuation – sequence of acts

Bakeman and Gottman's (1997) approach to analysing the sequential data, which explores behaviour resulting from events as they develop over time, was employed. Notably, no previous study analysed the sequential data during a flood evacuation, representing an additional contribution to this study. Sequential analysis is an efficient approach to identifying behavioural patterns, which offers valuable insights for emergency responders to organise people's responses to a flood evacuation event. It can detect and uncover people's perceptions and decision-making processes under risky situations.

It is essential to compute the transitional probabilities to identify the most probable behavioural pattern in a flood evacuation. Transitional probability is considered one kind of conditional probability, describing the probability of the target event B occurring immediately after event A (Bakeman & Gottman, 1997). Results show that the most probable behavioural pattern in a flood evacuation would include communicating with household members (A), collecting belongings and emergency supplies (D), getting into the car (B), and then heading to the safe point (H), as illustrated in Figure 6.

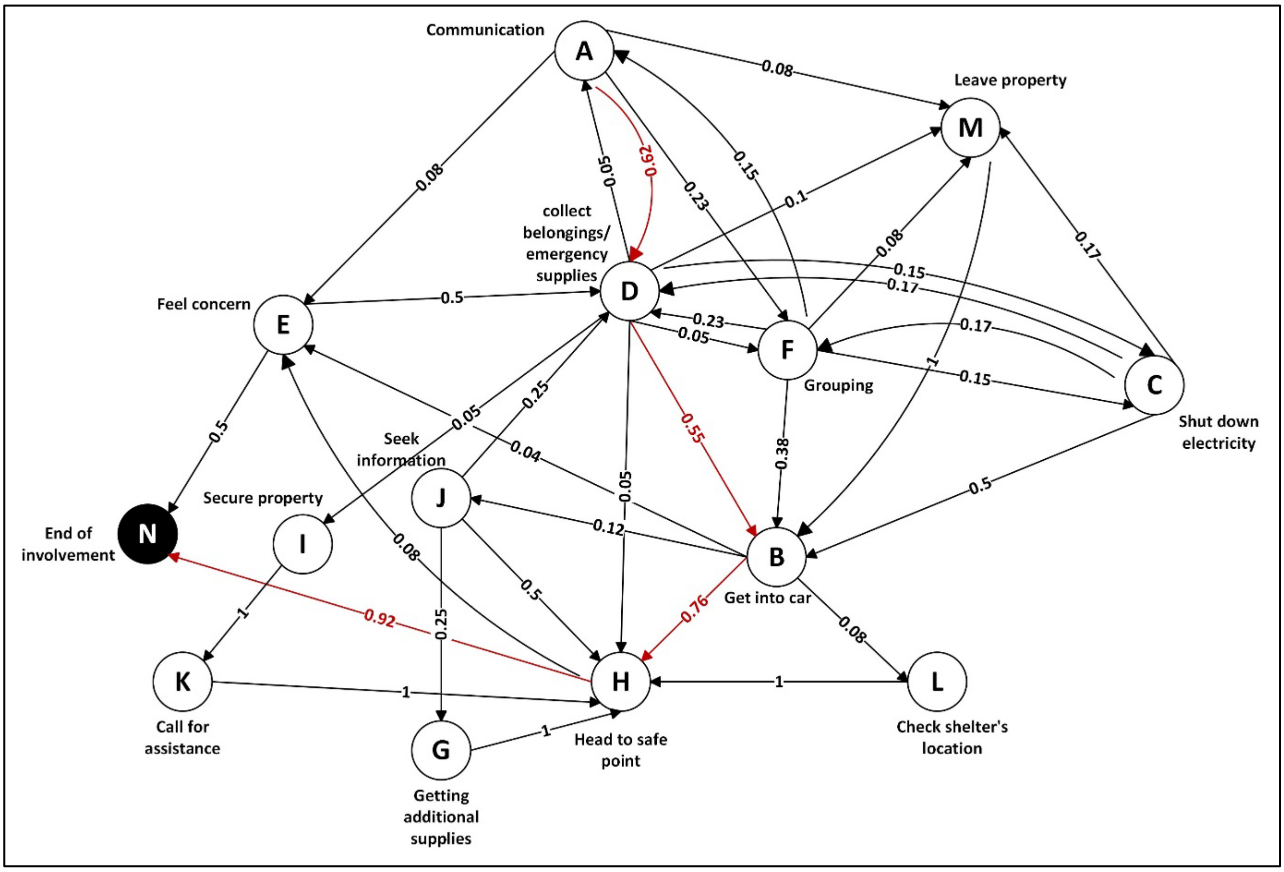


Figure 6: State transition diagram

Conclusions

This research has established the likely compliance behaviour concerning evacuation time and route choice, under different influences based on the PADM framework. Social cues, specifically noticing others complying, were the most effective factor in reducing evacuation delay time and enhancing compliance with evacuation orders. This was followed by environmental cues, specifically heavy rainfall and rising flood water. However, without any routing instructions or specific cues, evacuees ranked familiar and shortest (by distance) routes as more preferable than those taken by neighbours. No significant differences were seen for least congested, on high ground or recommended by Google Maps routes. This work highlighted the need for further research to study drivers' compliance, addressing route choice under multiple information sources and message content influences while travelling.

References

Aldahlawi, R. Y., Akbari, V. & Lawson, G., 2024. A systematic review of methodologies for human behavior modelling and routing optimization in large-scale evacuation planning. *International Journal of Disaster Risk Reduction,* Volume 110, p. 104638.

Bakeman, R. & Gottman, J. M., 1997. *Observing Interaction: An Introduction to Sequential Analysis.* Second Edition ed. New York: Cambridge University Press.

Bakhshian, E. & Martinez-Pastor, B., 2023. Evaluating human behaviour during a disaster evacuation process: A literature review. *Journal Of Traffic And Transportation Engineering,* 10(4), pp. 485-507.

Bian, R. et al., 2023. Modeling Evacuees' Intended Responses to a Phased Hurricane Evacuation Order. *Applied Sciences,* 13(8), p. 13085194.

Chiu, Y.-C. & Mirchandani, P. B., 2008. Online Behavior-Robust Feedback Information Routing Strategy for Mass Evacuation. *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS,* 9(2).

Cova, T. J. & Johnson, J. P., 2003. A network flow model for lane-based evacuation routing. *Transportation Research Part A: Policy and Practice,* 37(7), pp. 579-604.

Daudé, É. et al., 2019. *ESCAPE: Exploring by Simulation Cities Awareness on Population Evacuation.* Valencia, ISCRAM Conference.

Fu, H., Pel, A. J. & Hoogendoorn, S. P., 2015. Optimization of Evacuation Traffic Management With Intersection Control Constraints. *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS,* 16(1).

Lawson, G., 2011. *Predicting human behaviour in emergencies,* Nottingham: PhD Thesis, University of Nottingham.

Lindell, M. K., Arlikatti, S. & Huang, S.-K., 2019. Immediate behavioral response to the June 17, 2013 flash floods in Uttarakhand, North India. *International Journal of Disaster Risk Reduction,* Volume 34, pp. 129-146.

Lindell, M. K. & Perry, R. W., 2012. The Protective Action Decision Model: Theoretical Modifications and Additional Evidence. *Risk Analysis,* 32(4).

Lindell, M. K. & Perry, R. W., 2012. The Protective Action Decision Model: Theoretical Modifications and Additional Evidence. *Risk Analysis,* 32(4).

Pel, A. J., Hoogendoorn, S. P. & Bliemer, M. C., 2010. Evacuation modeling including traveler information and compliance behavior. *Procedia Engineering,* Volume 3, pp. 101-111.

Saha, S. K. & James, H., 2017. Reasons for non-compliance with cyclone evacuation orders in Bangladesh. *International Journal of Disaster Risk Reduction,* Volume 21, p. 196–204.

Wang, Z. et al., 2020. Analysis of Flood Evacuation Process in Vulnerable Community with Mutual Aid Mechanism: An Agent-Based Simulation Framework. *Int. J. Environ. Res. Public Health,* Volume 17.

Wilson, J. R. & Sharples, S., 2015. *EVALUATION OF HUMAN WORK.* Fourth ed. s.l.:Taylor & Francis Group.

Wu, H.-C., Lindell, M. K. & Prater, C. S., 2012. Logistics of hurricane evacuation in Hurricanes Katrina and Rita. *Transportation Research Part F: Traffic Psychology and Behaviour,* 15(4), pp. 445-461.

Prospective Cohort Study on Paramedic Fatigue: Impact of Workload and Shift Schedule

Amin Yazdani & Marcus Yung

Canadian Institute for Safety, Wellness, and Performance, Conestoga College, Canada

SUMMARY

Paramedics experience significant levels of fatigue that may affect their health and safety and the safety of the communities they serve. Through a 1-year prospective cohort study, combining ActiGraph watch data and ambulance call reports, we found possible relationships between workload (i.e., call volume), work arrangement (i.e., shift schedule), and sleep quality and duration.

KEYWORDS

Ambulance Call Reports, Sleep, Rural Paramedic Service

Introduction

Paramedics are at a high-risk for fatigue. Fatigue is a multidimensional construct and can lead to performance decrements which may endanger not only the health and safety of responders but also the safety of the public they serve (Yung et al., 2021). In a survey study of Ontario paramedic services, 55% of paramedic personnel reported being fatigued at work; fatigued paramedics were twice as likely to report injuries, three times more likely to engage in safety-compromising behaviours, and 1.5 times more likely to commit a medical error (Donnelly et al., 2020).

Fatigue may be attributed to work tasks (e.g., workload related to physical, cognitive, or psychosocial demands) and work arrangement (e.g., shift schedules, etc.). However, few studies have explored the relationships between work tasks/arrangement and sleep-related paramedic fatigue. This study explored these exposure-response relationships through a one-year prospective cohort study, gathering and analysing primary data (actigraphy sleep and activity measurements) and secondary data (administrative and daily ambulance call reports - ACR).

Method

Sixty-three primary and advanced care paramedics, from a large urban and large rural paramedic service, were recruited to participate in a prospective cohort study. Data was collected at three time periods, four months apart. During each collection period, participants wore a wrist-mounted ActiGraph device to monitor activity and sleep quality and duration, for 28 days, resulting in 84 days of actigraphy information from each participant. Participants also completed a sleep diary log to verify actigraphy data. For each participant, we gathered their administrative data to ascertain their shift schedule. We also analysed ACR data, which document daily events including workload factors (e.g., call volume, clinical procedures, total call time, transfer of care wait time, total drive time, and number of procedures). We estimated the daily emotional, physical, and mental demands of their calls using an "intervention demand matrix" (IDM), designed for this study, by retrospectively assigning workload demands to their performed clinical procedures.

For this presentation, we detail the development of the IDM and report on relationships between fatigue risk factors (workload- and work arrangement-related) and fatigue outcomes, determined by

ActiGraph sleep metrics (sleep efficiency, total sleep time, number of awakenings, awakening length, movement index, and fragmentation index). After linking fatigue risk factor data to actigraphy data, we performed logistic regression. We defined a "fatigue case" as a sleep event that met the following criteria: (1) ≤ 85% sleep efficiency, and (2) ≥ 6 average awakenings per night, and (3) >5 minutes average awakening length. We designated age, sex, BMI, job tenure, and chronotype as covariates in our models. We report preliminary univariate model results, both crude and adjusted.

Preliminary Findings

The IDM was developed, based on the NASA Task Load Index (TLX), to estimate physical, mental, and emotional demands of clinical procedures. Demand estimates were obtained for 64 commonly used procedures from frontline paramedics and management. We calculated winsorized means for each of the 64 procedures and assigned these demand estimates for each paramedic call. We calculated an average demand score (i.e., frequency weighted) and cumulative call demand score for physical, mental, and emotional demand dimensions.

We found statistically significant associations between workload factors and adverse sleep-related fatigue. Interestingly, we observed a u-shaped trend between call volume and fatigue, where paramedics with 2-3 calls per shift had 65% lower odds of being a fatigue case than paramedics with no calls per shift. We found a relationship between a one unit increase in the average mental, physical, and emotional demands of the intervention per shift and the risk of being a fatigue case, from 3-5% according to crude and adjusted models. We did not find a statistically significant relationship between shift type and sleep-related fatigue.

Table 1: Relationships between workload and work arrangement fatigue risk factors and fatigue case status.

Fatigue Risk Factor	Non-Cases (n)	Cases (n)	Crude OR (95%CI)	p-value	Adjusted* OR (95%CI)	p-value
Workload						
Call Volume						
No Calls (Ref)	282	30	1.00		1.00	
Low # of Calls (1 call)	89	9	0.95 (0.43 - 2.08)	0.90	1.15 (0.43 - 3.05)	0.79
Moderate # of Calls (2-3 calls)	135	5	**0.35 (0.13 - 0.92)**	0.03	0.76 (0.26 - 2.20)	0.61
High # of Calls (4+ calls)	16	3	1.76 (0.49 - 6.40)	0.39	**3.76 (0.88 - 15.97)**	0.07
Total Drive Time	240	17	0.99 (0.99 -1.01)	0.89	1.00 (0.99 - 1.01)	0.34
Transfer of Care Wait Time	240	17	1.00 (0.99 - 1.011)	0.31	**1.01 (1.00 - 1.02)**	0.02
Total Call Time	240	17	1.00 (1.00 - 1.004)	0.48	**1.01 (1.00 - 1.01)**	0.04
# of Interventions/Procedures Per Shift	240	17	**1.04 (1.01 - 1.07)**	0.01	**1.04 (1.01 - 1.08)**	0.02
Avg. Mental Demand of Interventions/Procedures Per Shift (FreqWt)	240	17	**1.03 (1.00 -1.10)**	0.05	1.02 (0.99 - 1.06)	0.09
Avg. Physical Demand of Interventions/Procedures Per Shift (FreqWt)	240	17	**1.05 (1.00 - 1.10)**	0.04	**1.05 (1.00 - 1.10)**	0.05
Avg. Frustration/Emotional Demand of Interventions/Procedures Per Shift (FreqWt)	240	17	**1.04 (1.00 - 1.09)**	0.05	1.03 (0.99 - 1.08)	0.18
Cumulative Mental Demand of Interventions/Procedures Per Shift	240	17	1.00 (0.99 - 1.01)	0.06	1.00 (1.00 - 1.01)	0.25
Cumulative Physical Demand of Interventions/Procedures Per Shift	240	17	1.01 (0.99 - 1.01)	0.06	1.01 (0.99 - 1.01)	0.15
Cumulative Frustration/Emotional Demand of Interventions/Procedures Per Shift	240	17	**1.01 (1.00 - 1.01)**	0.05	1.00 (0.99 - 1.01)	0.25
Work Arrangement						
Shift Type						
Rotating (Ref)	777	90	1.00		1.00	
Non-Rotating (incl. straight days/nights)	349	29	0.72 (0.46 - 1.11)	0.14	1.66 (0.70 - 3.91)	0.25

*Adjusted for Age, Sex, BMI, Job Tenure, Chronotype **Call volume includes zero call shifts. All other variables exclude zero calls.

Key Takeaways

- Mental, physical, and emotional workload demonstrated significant associations with adverse sleep-related fatigue.
- There may be a u-shaped trend between number of calls (call volume) and adverse sleep-related fatigue.
- Research findings may support better evidence-informed decision-making in designing optimal work schedules that accommodates paramedic operations in the safest manner possible.

References

Donnelly, E.A., Bradford, P., Davis, M., …., & Chandra Pichika, S. (2020). What influences safety in paramedicine? Understanding the impact of stress and fatigue on safety outcomes. *Journal of the American College of Emergency Physicians Open, 1*(4), 460-473.

Yung, M., Du, B., Gruber, J., & Yazdani, A. (2021). Developing a Canadian fatigue risk management standard for first responders: Defining the scope. *Safety Science*, 134, 105044.

Assessing the effectiveness of virtual reality tasks as stress-inducing environments

Debora Colodete[1,3], Bernard Costa[2,3], Marcus Baldo[2,3]

[1]University of South Florida, USA,[2]University of Sao Paulo, Brazil, [3]Cognittiv, Brazil

SUMMARY

Due to safety concerns, the impact of stress on cognitive readiness cannot be assessed in real-time hazardous work scenarios through many psychophysiological methods. This study aimed to evaluate the validity of a virtual reality (VR) simulation designed to replicate a critical situation, eliciting a significant level of stress. This approach enables the analysis of behaviours that may mirror those exhibited by workers in high-risk environments such as oil platforms, deep-sea diving, or bomb disposal units.

KEYWORDS

Cognitive readiness, stress, virtual reality, hazardous environments.

Background

The human brain's complexity allows for learning, task performance, and adaptive responses to stimuli. Cognitive readiness denotes a state in which alertness and mental preparation reach a level conducive to optimal performance (Fletcher, 2004; Crameri, 2019; O'Neil, 2013). This readiness becomes critical in complex and unpredictable environments. However, external factors like pressure and stress can impair cognitive readiness, posing long-term challenges in high-risk sectors, such as oil platforms, deep diving, or bomb disposal units, where workers often operate under adverse and dangerous conditions (Lafond, 2012; von Rosemberg, 2019; Chowdhury, 2025).

Innovative technologies capable of monitoring cognitive readiness through accessible, simple, and efficient tasks could help reduce work-related accidents in hazardous scenarios. However, for safety reasons, psychophysiological measures, such as electroencephalographic signals and electrodermal potentials, which provide valuable insights for developing such technologies, cannot be collected during real-time operations (Chowdhury, 2025). To address this experimental limitation, we evaluated the effectiveness of a virtual reality simulation involving a stress-inducing task, designed to emulate essential sensorimotor attributes of procedures performed in dangerous situations.

Materials and Methods

Fifty adult volunteers were initially recruited from among university students. However, six participants were excluded from the analysis due to corrupted or missing data. The final sample comprised 44 participants (21 males and 23 females), aged 18 to 30 years.

All participants completed two tasks:

(i) Bomb deactivation simulation**:** In this VR task, participants played the game *Keep Talking and Nobody Explodes*, having to dismantle a bomb under a noisy environment, increasingly complex instructions, and time pressure. This game was selected for its intrinsic potential to elicit acute stress and anxiety.

(ii) Affective image presentation**:** After the VR task, participants viewed a five-minute sequence of either positive (P) or negative (N) images selected from the International Affective Picture System (IAPS; Lang, 2005) on a computer screen. Participants were divided into two groups (P and N) according to the nature of the images presented. They were instructed to observe the images without any additional task.

The State-Trait Anxiety Inventory (STAI) was administered in its two forms. The State form (STAI-S) was applied at three key points to assess overall anxiety: at the start of the experiment (pre-test), at the conclusion of the VR task (post-game), and immediately after the presentation of the affective images (post-images). The Trait form (STAI-T), along with the Visual Analogue Mood Scale (VAMS), was administered following each task (post-game and post-images) to evaluate levels of tension, anxiety, and nervousness.

Electrodermal potentials (galvanic skin response, GSR), heart rate (HR), and electroencephalographic activity (EEG) were recorded from all participants during the behavioural procedures. The data derived from these measures are still under analysis and will not be presented in this report.

Results

This preliminary study focused on validating the stress-inducing VR environment by analysing the impact of the bomb deactivation task and the affective image sequences on participants' mood states.

Analysis of the STAI-S data (Figure 1) using a 2x3 two-way mixed-design ANOVA, with Group (positive vs. negative images) as a between-groups factor and Time (pre-test, post-game, and post-images) as a within-groups factor, revealed a highly significant main effect of Time ($F[84,2] = 20.33$, $p < 0.0001$, partial eta-squared = 0.326). A marginal interaction between the two factors ($F[84,2] = 2.548$, $p = 0.084$, partial eta-squared = 0.057) aligned with a post-hoc Holm's test, indicating no statistically significant differences between time points for the positive image group ($p > 0.095$). Conversely, a significant difference ($p < 0.001$) was observed between the pre-test and both post-game and post-images scores for the negative image group, with no significant difference between the latter two ($p > 0.99$).

A pairwise comparison (Holm's test) between the levels of the factor Time across the combined groups (P and N) confirmed the highly significant difference between pre-test and both post-game and post-images scores ($t > 4.829$, $p < 0.0001$), also with no significant difference between the latter two ($p = 0.957$).

VAMS data yielded similar findings. A 2x2 mixed-design ANOVA revealed a marginal interaction between factors ($F[41,1] = 2.843$, $p = 0.099$, partial eta-squared = 0.065), consistent with a non-significant decrease in anxiety scores only for the positive image group.

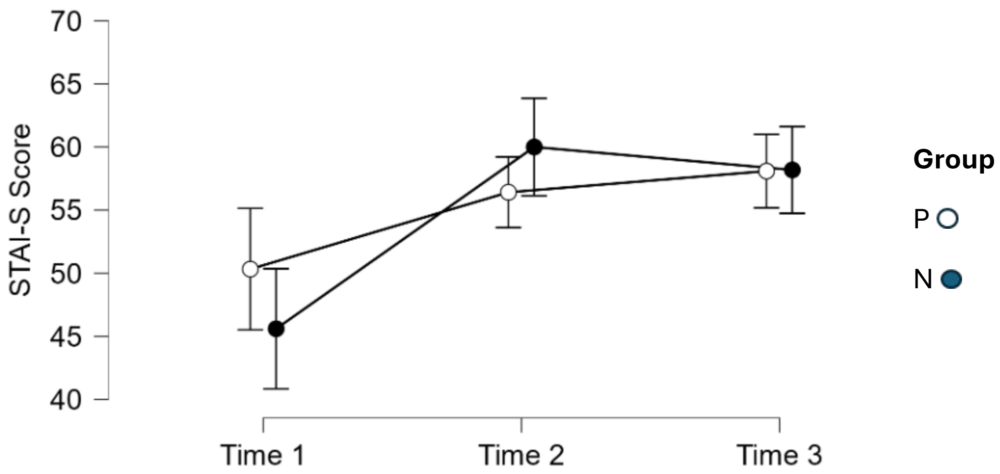


Figure 1: Mean scores from the State-Trait Anxiety Inventory scale (STAI-S). The lines show the STAI-S scores at three moments (Time 1 to 3 indicate pre-test, post-game, and post-images points in time) for both participant groups (P and N code the groups exposed to positive or negative images, respectively). Error bars denote 95% confidence intervals.

Conclusion

Our findings indicate that the VR-simulated task successfully induced a substantial level of anxiety, persisting during the presentation of negative images but diminishing towards normal levels with positive images. While the experimental procedure closely mimicked tasks performed by bomb disposal units, the physiological and emotional effects observed may generalise to other high-pressure, stressful environments, such as oil platforms, deep-sea diving and air traffic control towers, among others.

Further validation is required to confirm the task's relevance as a tool for investigating psychophysiological responses in real-world scenarios. However, these preliminary results suggest that VR-based simulations offer promising avenues for studying stress and cognitive readiness in settings where real-time data collection is unfeasible.

References

Chowdhury, T. I., Vargo, A., Blakely, C., Tag, B. and Kise, K. (2025). arXiv:2501.03537 [cs.HC].

Crameri, L., Hettiarachchi, I., & Hanoun, S. (2019). A Review of Individual Operational Cognitive Readiness: Theory Development and Future Directions. Human Factors, 63(1), 66–87.

O'Neil, H. F., Lang, J., Perez, R. S., Escalante, D. & Fox, F. S. (2013). What Is Cognitive Readiness?

Fletcher, J. D. (2004). Cognitive Readiness: Preparing for the Unexpected. IDA Document D-3061 Log: H 06-000702.

Lafond, D., et al. (2012). Support Requirements for Cognitive Readiness in Complex Operations. J. of Cognitive Engineering and Decision Making, 6(4), 393-426.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2005). International Affective Picture System (IAPS) [Database record]. APA PsycTests.

von Rosemberg, J. (2019). Cognitive Appraisal and Stress Performance: The Threat/Challenge Matrix and Its Implications on Performance. Air Med J., 38(5):331-333.

Developing Human Factors guidance for introducing Artificial Intelligence into the energy sector

Robert Becker & Stirling Tyler

User Centric Design Ltd

SUMMARY

The project team updated guidance for the Energy Institute on how the introduction of Artificial Intelligence (AI) and highly automated systems will affect Human and Organisational Factors (HOF) within the energy industry. The project team conducted a review of several papers on HOF and AI to develop new guidance specific to the energy sector.

The new guidance focuses on different areas of human performance that AI and highly automated systems will influence. It also details the steps that should be taken during design and implementation of highly automated systems and AI using check sheets and screening tools.

KEYWORDS

Artificial Intelligence, Trust, Ethics, Human-Machine Teaming, Human Performance

Introduction

The Energy Institute is the professional body for the energy sector worldwide. In 2024, they requested that their guidance document, "Guidance on human and organisational aspects of implementing new technologies," be updated.

It was requested that the update should focus on the human and organisational factors surrounding AI and highly automated systems, with information from existing publications such as the CIEHF's white paper *Human Factors in Highly Automated Systems* (McLeod, 2022) being used to inform the guidance updates. While this technology could augment human performance in the energy sector, they pose unique challenges that warrant new guidance.

Method

The project team reviewed the original documentation and noted areas for updating with content about AI and highly automated systems insights. The team then read and reviewed relevant publications on HOF, highly automated systems and AI (see references for a complete list). Insights gained from these papers were collated and considered from the perspective of people working in the energy sector.

The project team then conducted workshops with Energy Institute stakeholders to present research findings and proposed changes to be implemented into the guidance document. These workshops included international representatives from different energy companies who gave feedback on which topics they would like to see and what would be most beneficial.

Following the workshops, the project team updated the guidance document to include content about AI and highly automated systems and their influence on human and organisational factors within the energy industry. Alongside the updated guidance, the project team created check sheets and HOF specialist input screening tools for readers. The check sheets guide readers in considering specific human performance factors related to introducing AI or highly automated systems. The HOF specialist input screening tool can determine if HOF specialists are required during AI/highly automated system development and implementation.

Themes

The main topics captured in the new guidance are explained in this section.

Transparency

Transparency refers to the ability of AI and highly automated systems to explain how they work to operators to foster trust, improve Situational Awareness (SA) and allow for effective human-machine teaming.

Transparency is important for human operators to build accurate mental models of the system they are working with, reducing the risk of over-reliance and 'automation surprises' (Sarter et al., 1997) - where operators are confused or surprised by system behaviour due to it being incongruous with their mental model.

Succinctly explaining how these systems work to human operators poses a significant challenge. Systems such as deep neural networks, which mimic human brain functions, make it hard to explain and predict how a model will react to different stimuli over time - these are systems that learn and adapt their behaviour continuously, meaning that behaviour becomes less predictable as time goes on.

To tackle these challenges, highly automated systems and AI need to be designed in a way so that they offer easy-to-understand explanations of their behaviour and confidence in their outputs. Human Factors intervention during design and validation of these systems can help to ensure that the Human-Machine Interface (HMI) is transparent and usable - e.g., by including users in the design process and carrying out heuristic reviews and performance testing during system validation.

Employee Resistance

Organisations should consider the significant risk of potentially demotivating staff, which could lead to decreased work efficiency (e.g., due to boredom). Fears surrounding job security, changes to workload and requirements for additional training may spur employee resistance to implementing such systems.

It is important to clearly communicate the intent associated with implementing these systems and regularly collect and understand employee attitudes and perspectives. This can help organisations develop implementation and communications strategies that allay any undue employee concerns and encourage acceptance of AI or highly automated systems.

Trust & Confidence

Operators' trust and confidence in AI and highly automated systems are closely tied to transparency, as operators should ideally understand how a system works in order to invest the appropriate amount of confidence or trust in it.

For systems with inherent uncertainty, such as predictive models, operators need an appropriate level of confidence to avoid over-reliance or unnecessary scepticism. Over-trust can lead to complacency and over-reliance on a system, even when incorrect. Under-trust can lead to system underuse, increasing operator workload as they take on tasks the system was designed to handle.

Systems should be designed to provide clear explanations of their processes and decisions, thus prompting the operator to invest the appropriate amount of trust in the system output. For example, if a system responsible for load prediction explains to operators that it is using a training data set from several years ago, the operator may put less confidence in the predictions the system makes and adjust accordingly.

Fictional Scenario: *The latest AI Grid Load Prediction System ('GLoPS') forecasts that a city will consume 2670 kWh over the next 24 hours. It informs Stanley, the human operator, that this prediction is based on 2024 average energy statistics. However, Stanley knows a new events arena is opening the following day (which GLoPS has not accounted for). Rather than dismissing the AI's output entirely, Stanley adjusts the predicted energy load to reflect increased demand. While Stanley has not dismissed the system's output, he has not over-relied on it either - thus showing appropriate confidence in its output.*

Ethics

AI and highly automated systems can develop biases that affect how they process information and create outputs, negatively impacting operators and consumers. One example is algorithmic bias, where the underlying reasoning used by the system inherently disadvantages a specific group of people.

For instance, Obermeyer et al. (2019) found that an algorithm used in American healthcare prioritised white patients over black patients for additional care. The algorithm assumed that patients who spent more on healthcare were sicker and required more support. However, in the United States (US), those who access healthcare more frequently tend to have the financial means to do so. As economic status and race are closely linked in the US, the system, therefore, unintentionally favoured wealthier white patients over poorer black patients.

Similarly, in the energy industry, investment in energy infrastructure has often favoured affluent or urban areas. When AI models use historical performance data, such as outage frequencies or maintenance records, they may inadvertently perpetuate these balances.

Adopting guidance such as the EU Commission's "ethics guidelines for trustworthy AI" into the design of AI systems can help shape organisational policy to promote ethical AI usage. Where possible, these guidelines should be translated into design principles to ensure that AI systems operate ethically. For example, if an AI model is designed to explicitly label how it has arrived at a conclusion, including providing sources (Endsley, 2023) and allowing users to query its logic (Sujan et al., 2021) ensures that a degree of transparency has been embedded into the system at the design level.

Risk Reduction

Energy data and grid behaviour are constantly changing with the increasing use of renewable energy sources. Predictive models must be robust against model drift—a situation where outdated data leads to inaccurate predictions—which is especially dangerous when managing real-time energy distribution.

Unlike traditional automation, where outputs are consistent and predictable, highly automated or AI systems introduce an element of uncertainty (IxDF, 2024) by changing their outputs based on prior experience and contextual awareness. Users should be made aware of this through training and be instructed to exercise discretion when using these systems to ensure an appropriate level of trust is adopted (see Trust & Confidence).

User errors may stem from over-reliance on inaccurate system outputs and committing to inappropriate actions as a result. This, combined with skill fade and startle responses from long periods of underload, can result in delayed and impaired human reactions during system failures (see Failure Conditions).

Risk assessments must consider the transparency of the highly automated or AI system, i.e. how effectively it can communicate its data sources and logic to the user - standards such as IEEE 7001-2021 provide actionable guidance on introducing suitably transparent autonomous systems (IEEE, 2021).

The ongoing maintenance needs of the algorithm must also be considered, for example, periodic updates to an AI model's database, to ensure that it remains valid during its use in live operations and that the risk of model drift is mitigated.

Workload

Highly automated or AI systems can significantly enhance operator capabilities and responsibilities by performing tasks and subtasks that traditionally require extensive knowledge or experience. This effectively lowers the threshold for skill, experience and knowledge required to carry out a specific task. The immediate benefit seen here is the de-concentration of workload from specialist operators to the broader team, removing process bottlenecks and improving efficiency.

However, providing operators with new tasks, even if supplemented with automation or AI, will affect their overall workload and change their role profile. The system should be designed so that users understand how to use the technology properly, spot inaccuracy and bias, and know when to refer to a suitably qualified human operator for support.

Failing to address these needs may lead to increased stress and fatigue, as well as reduced job satisfaction and overall wellbeing. As mentioned in the Transparency section, the system's design must accommodate the user by providing a clear, intuitive interface that is unambiguous in its presentation of data and addresses key user needs for the task being conducted; this is best achieved by including the end-user in the design process.

Failure Conditions

If AI and highly automated systems are used to manage critical energy infrastructure, then sudden failure could lead to widespread disruption and outages. AI and highly automated systems in the energy sector should ideally be designed to degrade gracefully, maintaining partial functionality and allowing for a controlled shutdown rather than a sudden and complete failure. For example, if a server fails, the system can redistribute its load to other servers and suspend low-priority processes. This should be paired with a comprehensive warnings, cautions and alerts philosophy to ensure users are notified of reduced system performance, allowing them to intervene accordingly.

Furthermore, organisations should identify and substantiate human-based safety claims that capture what the operator is required to do in failure conditions and the extent to which their input will mitigate the risk posed by system failure. Risks associated with automating operator tasks include skill fade and under-load – these factors combined will reduce the speed and efficacy of human intervention in the event of system failure.

Methods such as Failure Modes and Effects Analysis (FMEA) could also be used to determine how automation or AI may fail, the consequences of such failure and what demands would, therefore, be made of human operators under such conditions.

Human-Machine Teaming

Highly automated systems and AI can be used to augment existing teams, increase efficiency, and reduce workload. This may motivate organisations to reduce team sizes and augment operator performance with highly automated systems or AI. This approach must be planned carefully – as it may perpetuate a negative perception of highly automated or AI systems as 'stealing' jobs and generating new tasks and workload drivers for existing staff (see Employee Resistance). Furthermore, new roles required for the ongoing maintenance of highly automated or AI systems may increase the size of existing teams to include specialist operators, e.g. AI specialists, data scientists and ethics experts.

AI is more dynamic, adaptive and flexible than traditional computing. In some instances, suitably advanced models may appear to fill the role of a team member rather than a tool or work aid. Depending on the maturity and intended use of the AI model, the impact on team composition and dynamics may range from providing task augmentation to fully realised human-machine teams.

The energy sector relies on interdisciplinary teams that include roles such as maintenance engineers, IT specialists, and control room operators. AI and highly automated systems must be designed to be intuitive across these disciplines, facilitating clear communication and ensuring that all team members can access and understand the system they are working with. To accurately account for the risk posed by such changes, consideration should be given to SA across a team. For a team to have 'good' SA, there needs to be consideration for how non-human elements of the team (i.e. AI) can communicate with human elements and vice-versa. Task analysis can be used to derive the SA needs of each team member.

For example, if human operators neglect to update an AI model with the latest weather data, it may fail to produce accurate energy production predictions for a wind farm. Failing to update the system with relevant data and the failure of the system to communicate its operating parameters and the data it is using may result in degraded productivity, team performance and safety.

Allocation of function and the oversight between human operators and the system should be clearly defined and signposted to all members of a human-machine team. In many cases, AI may act as a system to give a second opinion to human operators or augment task performance. While human operators will still retain overall control and responsibility for their tasks, introducing AI may augment performance, reduce task completion time, and boost productivity and safety.

Check sheets and HOF Specialist Input Screening Tool

The new guidance also includes check sheets and a screening tool for readers to use.

Check Sheets

The check sheets can be used as step-by-step guides for any organisation considering the introduction of AI or highly-automated systems. Each sheet includes a prompt (e.g. *Have the effects of highly automated or AI enabled systems on team composition been identified?*); a list of actions and considerations (e.g. *Identify new roles required for operating the system*), and a list of applicable tools or techniques to use (e.g. *Task Analysis; User Workshops)*. There are multiple sheets which cover the topics discussed in this paper.

HOF Specialist Input Screening Tool

The HOF Specialist Input Screening Tool acts as a decision-making framework to determine when specialised Human Factors expertise is required.

The tool consists of a guide word column (e.g. *System Design*); an accompanying question (e.g. *Does the system require the design of a user interface for interacting with automated and AI systems?*), a tickbox conditions column (*Applies/Does Not Apply/Not Sure*) and an implications column indicating what capabilities and benefits HOF professionals can provide (e.g. *Human factors professionals can help ensure interfaces are designed that are intuitive, user-friendly, and support efficient human performance*).

Conclusions

AI and highly automated systems have the potential to bring significant benefits to the energy sector by providing human operators with a powerful, adaptable tool to aid decision-making and automate existing tasks. This paper has highlighted significant HOF challenges that need to be addressed by human factors professionals during the design, validation and implementation of this technology.

A common theme to many of the challenges posed by AI and highly automated systems is the concept of transparency. If the highly automated system or AI cannot clearly explain how it has arrived at a conclusion, then all other issues highlighted in this paper are amplified. Without transparency, operators may not be able to invest appropriate confidence in the system (leading to over- or under- trust); fail to spot algorithmic bias; struggle to adapt to human-machine teams; resist using the technology or be unable to cope with the system in the event of a critical failure.

Established methods used by Human Factors professionals, such as FMEA; heuristic reviews and user-centred design philosophy can be used to ensure that AI and highly automated systems are designed to be transparent, useful and safe. These methods were highlighted in the check sheets and HOF Specialist Input Screening Tool, providing readers with guidance when implementing AI or highly automated systems in their organisation.

References

Endsley, M.R (2023) HFE perspectives on Artificial Intelligence and the future of work. Available at https://www.youtube.com/watch?v=Uu32ky_Z4yE. Accessed June 2024.

European Commission website. Ethics guidelines for trustworthy AI. Available at https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai. Accessed June 2024.

Flemisch, F., Heesen, M., Hesse, T., Kelsch, J., Schieben, A., & Beller, J. (2011). Towards a dynamic balance between humans and automation: authority, ability, responsibility and control in shared and cooperative control situations. Cognition, Technology & Work, 14(1), 3–18

Institute of Electrical and Electronics Engineers (IEEE) 7001-2021, IEEE Standard for Transparency of Autonomous Systems. Available at https://standards.ieee.org/ieee/7001/6929/. Accessed June 2024.

IxDF website. AI Challenges and How You Can Overcome Them: How to Design for Trust. Available at https://www.interaction-design.org/literature/article/ai-challenges-and-how-you-can-overcome-them-how-to-design-for-trust. Accessed June 2024.

McLeod, R. (2022) Human Factors in Highly Automated Systems. Chartered Institute of Ergonomics and Human Factors. Available at https://ergonomics.org.uk/resource/human-factors-in-highly-automated-systems-white-paper.html. Accessed June 2024.

Microsoft website. Artificial Intelligence (AI) vs. Machine Learning (ML). Available at https://azure.microsoft.com/en-gb/resources/cloud-computing-dictionary/artificial-intelligence-vs-machine-learning. Accessed June 2024.

National Health Service AI Lab website, Moving from trust to appropriate confidence. Available at https://digital-transformation.hee.nhs.uk/building-a-digital-workforce/dart-ed/horizon-scanning/understanding-healthcare-workers-confidence-in-ai/executive-summary-and-report-overview/moving-from-trust-to-appropriate-confidence. Accessed June 2024.

Nielsen Norman Group website. CARE: Structure for Crafting AI Prompts. Available at https://www.nngroup.com/articles/careful-prompts/. Accessed June 2024.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453.

Sarter, N. B., Woods, D. D., & Billings, C. (1997). Automation surprises. Handbook of Human Factors and Ergonomics. Vol. 2, 1926-1943.
Stanton, N. A., Stewart, R., Harris, D., Houghton, R. J., Baber, C., McMaster, R., Salmon, P., Hoyle, G., Walker, G., Young, M. S., Linsell, M., Dymott, R., & Green, D. (2006). Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology. Ergonomics, 49(12–13), 1288–1311.
Sujan, M. (2021) Human Factors and Ergonomics in Healthcare AI. Available at https://ergonomics.org.uk/resource/human-factors-in-healthcare-ai.html. Accessed June 2024.

Inter-Organisation collaboration in Gas Distribution: A pathway to more consistent operations?

Harvey McIntosh[1] & Nikki Legg[2]

[1]SGN, [2]Cadent

SUMMARY

Four Gas Distribution Networks operate across Great Britain, distributing gas from the coastline to the gas meters. Although Human Factors has historically been applied across the industry, the individual organisations now employ Human Factors specialists in-house. This paper sets out how the industry came together to apply Safety Critical Task Analysis consistently across the industry. It discusses their methods, the challenges they faced and the lessons they learned along the way.

KEYWORDS

Gas Distribution, Safety Critical Task Analysis, Collaboration

Introduction

Four Gas Distribution Networks (GDNs) operate across Great Britain with their work divided by their geographical locations. GDNs are responsible for the infrastructure used to transport natural gas, biomethane, and shortly, hydrogen. They transport the gas from our coastline through to the meters in our homes. Additionally, GDNs provide an emergency service relating to gas escapes and related emergency scenarios, for example, carbon monoxide leaks.

In 2021, the GDN industry began recruiting in-house Human Factors specialists to support them with Human Factors Integration. Although each organisation has received Human Factors support before, this was the first time that the Human Factors discipline has been brought in-house. There is a huge breadth to the type of Human Factors activities completed, however, a main consideration for the gas distribution industry has been the application of Safety Critical Task Analysis (SCTA) to eliminate and/or mitigate the risk of human error to a level deemed as low as reasonably practicable.

SCTA can be defined as "the study of what a person is required to do, in terms of actions and mental processes, to achieve a goal, about Major Accident Hazard safety analysis" (Energy Institute, 2020). In the context of gas distribution, there are many instances where human performance could cause or contribute to major accident hazards or significant incidents and therefore it is important that these tasks are appropriately analysed to identify engineered and administrative controls which can be put in place to either eliminate or mitigate human error risk.

Inter-organisational collaboration

When the GDNs began the process of introducing SCTA to the industry's Safety Critical Tasks, it was proposed that industry collaboration should be introduced to create industry agreed SCTAs. It was hypothesised that a GDN SCTA collaboration would support the production of an industry Safety Critical Task Register and improve the consistency of operation across the GDN industry.

The inter-organisational sharing of results would provide insight into possible human errors and the controls put in place to eliminate and/or mitigate against these. To do this several methods were introduced.

An early requirement for effective collaboration and the ability to produce a Safety Critical Task Register was the definition of a joint approach to SCTA. This included agreeing on definitions for Safety Critical Tasks and the way in which they would be applied. Further, the industry needed to agree on the methods they would adopt for carrying out SCTA, from initial task screening through to completion and delivering actions.

The GDNs came together and agreed on the following definition for Safety Critical Tasks:

"A Safety Critical Task is one in which human performance could cause or contribute to a major accident hazard or significant incident or fail to prevent or reduce the effect of one."

To successfully work through a defined Safety Critical Task Register, the GDNs needed to follow the same steps to complete their SCTAs. By following a consistent approach GDNs could each work on separate analyses and then share their findings. GDNs could run through a check on each other's work, ensure it applied to their own organisation and then add it to their completed task register. Therefore, they agreed to combine guidance from several resources to create the following step-by-step method that they would all adopt.

Identify Safety Critical Tasks	Screen Safety Critical Tasks	Hierarchical Task Analysis	Human Error Analysis	Risk Reduction	Safety Critical Task Analysis Report	Action management/Review
Taken from a range of sources including organisational documentation, hazard assessments, risk assessements and Incident investigations	A two step process, the first taken from the Energy Institute (Energy Institute, 2020). And the second step taken from IChemE (Miguel, 2019)	Defining the task and breaking it down into steps	Analysing the steps to understand the potential for these steps to go wrong	Work to reduce the risk of the step going wrong or reduce the consequences if it does go wrong	Create a report summarising the findings of the safety critical task analysis	Record actions, who's responsible and the date of completion

Figure 1: Agreed steps for the GDN SCTA process.

To produce the industry Safety Critical Task Register each GDN worked independently, with their subject matter experts, to create a Safety Critical Task Register applicable to their own organisation. These registers were created from a review of procedural documentation, a review of the safety case and a review of associated assessments (bowties, LOPAs etc.), along with findings and recommendations from incident investigations. Early comparison of task registers highlighted a large degree of similarity, but also some differences in the tasks captured. The GDNs were once one company, so it is logical that there are some tasks that are carried out universally and some task differences which are driven by the specific geographical and operational context of today's organisational structure.

Collaboration to determine the crossover between the individual task lists would enable the GDNs to become more efficient and be able to share results as SCTA progressed. Therefore, GDN Human Factor specialists collaborated as part of a two-day workshop to synthesise the similar, industry-wide tasks into a single industry-wide Safety Critical Task Register. The Safety Critical Tasks listed on each register were cross-referenced across each organisational task register to assess their applicability to either the industry as a whole or just a single organisation. Those that applied across the industry were included on the industry-wide Safety Critical Task Register. In total 52 tasks have been included on the GDN industry Safety Critical Task Register to date. It is these tasks which have been agreed for sharing and collaboration as each GDN works through the task register.

As SCTAs are completed, the GDNs share their outcomes with each other. Analyses are then reviewed against the other GDNs procedures for completing the task, alongside engineers that complete the tasks. Where the tasks differ across organisations the SCTAs need to be adapted; where they are the same analysis can be carried forward.

The GDNs have committed to carrying out at least four SCTAs each per year. With collaboration and sharing of the task analyses across the GDNs, this should mean that 16 SCTAs are completed across the industry, each year, which is over and above what a single organisation could complete, alongside competing Human Factors integration requirements.

Lessons Learnt

Although the GDNs are still at the start of their collaborative journey several benefits of the defined inter-organisational collaboration approach have already been realised. It has been possible to agree on both an industry definition and method of working in relation to SCTA. These are captured as part of an industry standard against which all GDNs conform. As each GDN is responsible for completing their own SCTA as they relate to the specific operational context, the use of a consistent methodology allows for effective benchmarking of findings, recommendations, and consequent safety and efficiency performance across the industry.

A second benefit of an industry agreed standard is that a formal mechanism for inter-organisational learning can be established. As screenings and analyses are completed in the same way using the same data resources, tasks and results can be shared and compared across organisations. It is not the case that results can be directly copied and pasted because the operational context is not identical in each organisation. However key safety-related findings can be shared for awareness and to influence the completed analysis. This promotes consistency across the industry and has the potential to increase the speed with which SCTA can be conducted. To support sharing of information, regular inter-organisation review meetings are held to discuss our analysis and identify areas and opportunities to share learning. This is formalised by the presence of a governance group who can ratify findings and recommendations.

It is noteworthy that the gas distribution industry is at the start of its HFI journey and therefore embedded HF Specialists work either independently or in small teams of two people. This can make it difficult to collaborate on HF projects and to achieve the level of review required for assessment relating to safety critical tasks. By establishing a joint approach to SCTA the industry has provided a platform for HF Specialists to create an inter-organisational HF team who can collaborate on a single industry-wide goal. This also provides HF Specialists with the benefits of being able to work together, constructively challenging one another’s ideas and thinking, and ultimately contributing to the safe and effective integration of HF into the gas distribution industry. In a discipline which can often feel isolating it is of great value to have the support of a team who speak the same language, share the same concerns, and can provide constructive input to overcoming the integration challenges which can be faced by HF Specialists supporting HFI in any business or industry.

Despite the positive lessons learnt to date, the application of a consistent approach across the industry has not been without its difficulties and there have been several challenges to overcome. Most noticeably, differences in organisational context and operational processes and procedures have highlighted that although a shared safety critical task register could be created, there is also a need for individual organisational safety critical task registers to be kept. This is because although many tasks are shared, each GDN also carries out safety critical tasks which are unique to them. For example, only one GDN completes operations relating to Liquified Natural Gas and Liquified Petroleum Gas; another GDN is the only one to manage the storage of large volumes of gas in an underground salt cavern. Failure of the human tasks associated with management of these activities could foreseeably contribute either directly or indirectly to a MAH and therefore it is important that SCTA is completed for these tasks.

As the industry has continued to work through analysis of their shared safety critical task register it has also become apparent that specific organisational content has an impact on potential errors identified during human error analysis and the effectiveness of controls put in place to eliminate or mitigate these errors. Therefore, it has not been and will not be possible for GDNs to copy and paste one another's work, reducing the impact of any efficiency gains hypothesised as part of the initial approach definition. To address this, individual GDNs have put in place additional checks to confirm that any analysis copied from another GDN is relevant and applicable to them. As such, GDNs are now using one another's SCTA to inform their own analysis rather than to act as their analysis. This allows adjustments to be made to analyses to reflect organisational differences, whilst still promoting a level of consistency by allowing findings and controls from other GDNs' inform the analysis process.

Although the four GB GDNs were derived from the same organisation, National Grid, further confounding the ability to make full use of one another's SCTA is the lack of a common language. This has hindered both the creation of an industry-wide safety critical task register and the consequent analysis of identified tasks. Disparity if driven by the difference in technical language and equipment used in each organisation. To address this, where necessary, safety critical tasks were defined using neutral language with specific notes added for individual organisations.

A pathway to more consistent operation?

We have already achieved several milestones in the proposed journey. Establishing an inter-organisational working group has allowed the industry to agree on a definition and method for SCTA. Using this methodology the four GDNs have been able to collaborate and produce an industry-wide Safety Critical Task Register. The Safety Critical Task Register captures Safety Critical Tasks found across the industry and maps these to each of the GDNs, highlighting opportunities for sharing and learning. The Safety Critical Task Register also highlights to GDNs where their operations may differ from those found across the rest of the industry and therefore where additional attention should be directed.

The industry is also successfully piloting a mechanism for sharing analyses and using these to support replication in each of our respective organisations. Although we are still early in this process it is hoped that this will support the adoption of consistent risk controls across the industry and speed up the process of SCTA by reducing the level of workload required. This is an especially important benefit in an industry and wider external environment which faces Human Factors resource limitations.

Although the gas distribution industry has advanced to incorporate a consistent method of risk assessment relating to safety critical tasks, it is too early to provide evidence to either support or refute the hypothesis that inter-organisational collaboration on SCTA will lead to a more consistent operation. More time is required to observe and evaluate the way in which different organisations approach the integration of recommendations within their specific organisational context.

Conclusion

This paper sets out to explore whether true inter-organisational collaboration can occur within this industry, by exploring their operational similarities and differences and the SCTA challenges that they face. Despite facing several obstacles early indications suggest that it is possible for multiple organisations to collaborate on and support SCTA for an industry. Whether this collaboration is an effective vehicle to promote industry consistency and improve safety is still to be determined. We are at the start of this inter-organisational collaboration journey so constant evaluation of the approach and effectiveness will be essential to determine whether inter-organisational collaboration is a pathway to more consistent operations.

References

Energy Institute. (2020, February). *Guidance on human factors safety critical task analysis.* Retrieved from Energy institute publications: https://publishing.energyinst.org/topics/human-and-organisational-factors/risk-management/guidance-on-human-factors-safety-critical-task-analysis2

Miguez, R; Roels, R; Cameron, J;, (2019). *Safety Critical tasks and Management of Human Error.* Retrieved from IChemE publications. https://www.icheme.org/media/19458/hazards-29-poster-10.pdf

Building a national, systemic network to advance quality and patient safety research

Jill Poots[1], Sam Cromie[1], Gemma Moore[2], & Dr Orla Healy[2]

[1]Trinity College Dublin, [2]Health Service Executive

SUMMARY

Capacity for quality and patient safety research can be limited by challenges such as: lack of time, funding, knowledge, and motivation. Research networks can help overcome these barriers by pooling knowledge, skills, and resources, and bolstering collaboration. This paper describes the journey so far to build a national QPS research network aiming to bring together members from across academia and the health system – from patients – to workers – to leaders – to encourage development and adoption of evidence-based approaches to system improvement.

KEYWORDS

Systems approaches, networks, healthcare

Introduction

As of 2021, the seminal paper, 'To Err is Human' (Institute of Medicine, 2000), which called for a systems approach to patient safety had been cited approximately 20000 times (Pierre et al., 2021). Using this as a marker for systems approaches to patient safety, it may appear that most research has been carried out in the USA with other countries lagging significantly behind. This may be true of the Irish healthcare system, with a scoping review of Irish patient safety research suggesting recently there is a 'modest' amount of patient safety research (O'Connor et al., 2023) and barriers identified to adopting human factors approaches such as lack of shared language, expertise, and a competency framework (Sharafkhani et al., 2024). Moreover, anecdotal evidence suggests this research has often been carried out in institutional or geographical 'siloes', which can lead to duplication, research waste, and lack of implementation of evidence-based practice, to improve patient safety.

Improving patient safety in the Irish healthcare context is imperative, based on evidence from two studies suggesting the rate of adverse events in hospitals showed no decline between 2009 and 2015 (Rafter et al., 2016; Connolly et al., 2021). To address this, a national system-wide network has been established aimed at bringing together stakeholders across the healthcare system, from patients, to frontline staff, to healthcare leaders, and academics, to identify priorities for quality and patient safety (QPS) research and improvement work, and contribute to better, safer care. Such research networks may promote collaboration between different institutions and stakeholders and increase knowledge transmission (Adams, 2012) and trustworthiness in research (Nyirenda et al., 2020). This paper outlines the journey and priority QPS research areas identified so far, which include sociotechnical systems design, and implementation science.

Structure and Scope of the network

The Evidence-Based QUality Improvement and Patient Safety Research Network (EQUIPS) was established in 2024 and is co-funded by the Health Research Board (HRB) and Health Service Executive (HSE) National Quality and Patient Safety Directorate (NQPSD). The network itself adopts a systems and co-design approach: 40 grant co-applicants and collaborators, representing multiple academic institutions, healthcare departments, public bodies, and patient advocacy groups came together to apply for the funding and develop a programme of work.

The network's work programme is divided into three activity 'strands'. Each strand and the tasks within them have a leader representing each of the stakeholder groups (i.e., patient partners, healthcare professionals and academics). The Enabling strand puts in place activities and resources to strengthen capacity for and capability to do QPS research, build a vibrant, sustainable community, and support knowledge translation into everyday practice. Through systematic review methods and stakeholder consultations, the 'Understanding and Informing' strand aims to ascertain priorities and strategies for QPS research in Ireland, develop shared definitions of quality and safety, and identify barriers and facilitators to participation in QPS research. This strand also hosts a social network analysis to analyse the growth of the network and the factors supporting growth and knowledge dissemination. The 'Focusing' strand is the research engine of the network, which is divided into research 'clusters' aiming to address research gaps in identified priority areas. The first priorities identified are Sociotechnical Systems Design and Implementation Science.

Progress and challenges

By January 2025, the EQUIPS Research Network grew from the initial 40 co-applicants and collaborators to a community of over 200 members in Ireland and further afield. To date, the EQUIPS Research Network has held several successful events, including an in-person event in Dublin which facilitated networking, hosted a keynote speaker who delivered an introduction to systems thinking, and stimulated conversations regarding the barriers and facilitators to patient safety. Progress has been made on its work programme, including two literature reviews (the findings from which should be available by the time of the conference), a webinar series on 'the practicalities of QPS research', and development of a database of QPS research and improvement projects. A social network analysis is planned to begin early in 2025 to provide a snapshot of the system, including thought leaders, knowledge brokers, and peripheral members, as well as provide an indication of knowledge exchange.

Building such a network is not without its challenges. Facilitating collaboration and equal participation between such varied stakeholder groups can be difficult. To overcome this, the network is drawing on best practice guidance from other Irish research networks and setting up a 'ways of working' task force to develop a strategy for co-production and collaboration in genuine partnership in the network. To facilitate the attendance of frontline clinicians, which was an initial challenge, meetings have been shortened.

Conclusion

The EQUIPS Research Network aims to coordinate and accelerate quality and patient safety research in Ireland by bringing QPS researchers, knowledge users, those working in healthcare and patient partners together to improve care outcomes including safety and quality. By taking a systems approach to its development, engaging stakeholders at various levels of the system, it is hoped this aim will be met. This paper outlines initial successes and challenges, to support others to develop similar networks to bolster capacity for research and improvement work in healthcare.

References

Connolly, W., Rafter, N., Conroy, R. M., Stuart, C., Hickey, A., & Williams, D. J. (2021). The Irish National Adverse Event Study-2 (INAES-2): longitudinal trends in adverse event rates in the Irish healthcare system. *BMJ Quality & Safety*, *30*(7), 547-558.

O'Connor, P., O'Malley, R., Kaud, Y., Pierre, E. S., Dunne, R., Byrne, D., & Lydon, S. (2023). A scoping review of patient safety research carried out in the Republic of Ireland. *Irish Journal of Medical Science (1971-)*, *192*(1), 1-9.

Rafter, N., Hickey, A., Conroy, R. M., Condell, S., O'Connor, P., Vaughan, D., ... & Williams, D. J. (2017). The Irish National Adverse Events Study (INAES): the frequency and nature of adverse events in Irish hospitals—a retrospective record review study. *BMJ quality & safety*, *26*(2), 111-119.

Sharafkhani, M., Browne, M., Codd, M., O'Dea, A., Breen, D., Byrne, D., ... & Ward, M. E. A 5 year snapshot of education, research and publications about ergonomics in Irish Healthcare. In: Contemporary Ergonomics and Human Factors 2024. Eds. D Golightly, N Balfe & R Charles, CIEHF.

Developing an evidence-based safety performance framework for telephone triage

Jill Poots[1], Jim Morgan[2], Matteo Curcuruto[3], Stephen Elliott[4] & Andrew Catto[4]

[1]Trinity College Dublin, [2]Leeds Beckett University, [3]European University of Rome, [4]Integrated Care 24

SUMMARY

This paper outlines a research project aiming to develop a framework to support telephone triage organisations (e.g., NHS 111) investigating safety incidents or planning system changes. The framework, developed using systematic review and Delphi methods, identified factors at multiple levels of triage work systems. Notable gaps meriting further research and consideration by system leaders included the effect of the physical environment on system performance.

KEYWORDS

Telemedicine, primary care, patient safety

Introduction

Telephone triage refers to the assessment of callers' symptoms and signposting to an appropriate healthcare service for their needs. For years, it has been a popular service worldwide and usage increased following the social distancing guidance imposed during the COVID-19 pandemic. The NHS 111 – the telephone triage service available in England and Wales – receives on average, over one million calls per month (NHS, 2024). Though popular and widespread in use, telephone triage as a healthcare activity has received relatively little attention in research literature (Poots et al., 2024). Moreover, telephone triage imposes unique challenges for patient safety compared to face-to-face care, since callers are not physically co-present, placing a heightened onus on callers to accurately report their symptoms and – of particular relevance to NHS 111 – call handlers are not always clinically trained (Morgan & Muskett, 2020). Since many existing frameworks for learning from incidents are designed with secondary (hospital) care in mind, this can be problematic for learning from incidents. The aim of this research was to identify factors from earlier studies that influence patient safety in telephone triage and assess their generalisability to the NHS 111 system in England and Wales through expert consensus.

Methods

First, a systematic literature review and best-fit framework synthesis was undertaken to identify a) how safety has been measured previously in primary care telephone triage organisations and b) potential factors contributing to telephone triage safety. Following an initial bibliometric analysis of the same databases (Poots et al., 2024), PubMed and Scopus were searched for studies evaluating safety in primary care telephone triage organisations globally, yielding 4121 papers. These were screened for eligibility by two researchers (see Poots et al., 2022 for the full protocol), resulting in the inclusion of 96 studies. Using the Systems Engineering Initiative for Patient Safety (SEIPS) 2.0 model (Holden et al., 2013) as the "best-fit" framework, data from the studies were extracted in alignment with the SEIPS components.

To ascertain the generalisability of the review findings and identify additional factors, a modified Delphi study was conducted. An expert panel comprising 23 professionals from academia and healthcare (mostly from Integrated Urgent Care) was recruited through opportunity and snowball sampling. The experts stated their agreement with 138 factors identified from the review presented via online survey (consensus was defined as 70% agreement or disagreement) and offered additional factors not accounted for. This process continued until consensus was reached on all of the items (three rounds in total). Twenty-one experts participated in Round 2, and 16 in Round 3. To enhance framework usability, the agreed factors were arranged thematically using reflexive thematic analysis (Braun & Clarke, 2019). The framework was presented to the panel for feedback on its usefulness and practicality.

Results

Most evaluative research on telephone triage safety has been conducted primarily in USA, UK, and Australia. Safety was most often measured by authors as the appropriateness of the service offered to telephone triage patients. Most research studies focused on individual person factors (either the patient or triage professional; *n* = 86), followed by tools and technology factors (*n* = 35 papers). There were very few identified research papers (*n* = 6) which considered the effect of the physical environment on safety.

Generally, there was high agreement with the factors synthesised, and many suggestions. For example, in the first round of the modified Delphi study, 115 items were agreed for inclusion, and an additional 66 were suggested by the panel. Two more rounds followed, and the final list of factors totalled 186 after including and agreeing or disagreeing on new suggestions. To improve usability of the framework, these were arranged into themes such as, ‘technology design can affect safety’, ‘technology use can affect safety’. These were organised into a visual format which will be presented at the conference. Feedback was positive regarding the representativeness of the framework, with suggestions for accessibility of the framework design.

Discussion

With telemedicine and e-health use increasing, human factors can provide a useful lens through which to investigate these systems and understand how their design can affect system outcomes. The framework developed here highlights 186 unique, evidence-based factors in telephone triage, validated with experts, which could support the investigation of safety incidents and stimulate thought when planning for system changes. Further application and refinement are required to improve the usefulness of the framework in local systems.

The systematic review revealed a significant knowledge gap concerning the impact of the physical call centre environment on telephone triage performance, which differs markedly from hospital and clinical settings. Additionally, few studies examined interactions between different levels of the system, indicating a need for more holistic research approaches. While the review showed that most research has been conducted in Westernized countries, it is important to acknowledge a limitation of this study: only English-language studies were included. Furthermore, as most experts consulted were familiar with telephone triage systems in England and Wales, the framework’s applicability in other countries has yet to be tested.

In conclusion, this study provides a robust framework to support telephone triage organizations, such as NHS 111, in improving patient safety through enhanced organizational learning (e.g. incident investigation) and more effective planning for system changes (e.g., procurement processes).

References

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health*, *11*(4), 589-597

Holden, R. J., Carayon, P., Gurses, A. P., Hoonakker, P., Hundt, A. S., Ozok, A. A., & Rivera-Rodriguez, A. J. (2013). SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. *Ergonomics*, *56*(11), 1669-1686.

Morgan, J. I., & Muskett, T. (2020). Interactional misalignment in the UK NHS 111 healthcare telephone triage service. *International journal of medical informatics*, *134*, 104030.

NHS England (2024). Integrated Urgent Care, England Aggregate Data Collection, September 2024. Accessed (26 Nov 24) at: https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2024/11/Statistical-Note-IUCADC-September-2024.pdf

Poots, J., Morgan, J. I., & Curcuruto, M. (2022). A Systematic Review and 'Best Fit' Framework Synthesis Exploring Contributory Factors to Safety in Primary Care Telephone Triage. [Protocol] PROSPERO 2022 CRD42022320933 Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022320933

Poots, J., Morgan, J., & Curcuruto, M. (2024). A Bibliometric Analysis of Telephone Triage Research to 2021 Using VOSviewer. *BioMed Research International*, *2024*(1), 5583853.

Developing healthcare safety investigator competencies for a consensus study

Sophie Hide and Rosemary Lim

Health Services Safety Investigations Body, UK

SUMMARY

There is a drive to professionalise healthcare safety investigators but there are currently no agreed core competencies. This paper describes the process of generating an initial set of competencies for healthcare safety investigators. The outputs formed the basis for a later consensus study.

KEYWORDS

Competencies, healthcare safety investigators

Introduction

Recent patient safety initiatives promote a systems approach to investigations (NHSE, 2022), but there is limited available resource from which to propose investigator competencies. Our wider study aimed to develop a competency framework for healthcare safety investigators in England. We report an early phase of the study in this paper: the development of an initial set of competencies for healthcare safety investigators, for later use in a consensus study.

Methods

Data source

We used existing interview data 'vignettes' from a project exploring the scope of a good and effective healthcare safety investigation. Interviews were conducted with 50 international and national safety specialists with varied expertise in contemporary healthcare safety investigations.

Analysis and competency development

Salient interview vignettes (n= 886), that varied in length (10-200 words), were analysed by two researchers with a varied background in healthcare, patient safety and human factors / ergonomics. For the first (approximate) third of interview vignettes (n=316), each researcher independently created 'competency descriptors' (i.e. describing a practice or quality of an investigator). These competency descriptors were discussed in depth to reach agreement. Thereafter, for the remaining vignettes (n=570), either researcher proposed competency descriptors for consideration by the other. Differing interpretations were resolved through discussions.

The competency descriptors (n=1214) were then grouped by meaning and application. These were informed by the relevant literature (e.g. Lester, 2014; Nixon and Braithwaite, 2018). Four distinct 'domains' were developed (see Table one), with subsidiary 'domain areas' (n=9) and 40 draft competencies (aggregates of the competency descriptors). The emphasis upon, and scope, of 'Personal qualities' and 'Effective and compassionate engagement' were noted for this role.

Results

Table one shows the draft competencies for healthcare safety investigators.

Table 1: Draft investigator competencies for use in the consensus study

Domain (n=4)	Domain area (n=9)	Draft competencies (n=40)
Personal qualities	Inherent characteristics	Is empathic
		Is inquisitive
		Is rigorous
	Professional identity	Demonstrates knowledge in healthcare safety investigation
		Invests in own professional development
		Supports development of knowledge in investigation practice
		Is perceived as a credible professional
		Promotes team-working
	Ethical practice	Demonstrates integrity
		Demonstrates diplomacy
		Reflects on the impact of their own expertise and experiences
		Is sensitive and responsive to the needs of investigation participants
		Seeks and offers peer review
		Demonstrates independence
		Maintains confidentiality
Investigation knowledge and skill application	Demonstrates core knowledge and understanding of:	The investigation context
		Complex healthcare systems
		Principles core to a healthcare safety investigation
		Principles of engagement
		Principles of just culture
	Investigation practice: Demonstrates applied skills as an investigator to:	Plan, prepare and review the design of an investigation
		Use systems-based investigation methodologies
		Use investigation methods and techniques
		Collect investigation data
		Analyse and interpret data
		Develop recommendations
		Report the investigation
		Assure quality of investigation integrity
Effective and compassionate engagement	Being supportive of people involved in an investigation	Demonstrates insight into the experiences of others
		Manages care of those involved in an investigation
	Meaningful engagement with people	Prepares for engagement
		Engages intentionally and genuinely with people
	Promotes engagement through interview practice	Prepares and supports interviewees
		Conducts the interview
	Writes for the audience	Tailors written materials for the user

<table>
<tr><td rowspan="3">Manages investigation lifecycles</td><td colspan="2">Identifies opportunities to learn from reported incidents</td><td>Manages data</td></tr>
<tr><td>Uses established investigation protocols</td><td colspan="2">Schedules work and manages time</td></tr>
<tr><td colspan="3">Monitors and measures the impact of investigation outputs</td></tr>
</table>

Key learning points from this process

To our knowledge, this draft competency framework is the first of its kind for healthcare safety investigators globally. The study demonstrates the rich outputs that can be gained from secondary data analysis. For later use in a consensus study, the outputs allow for review and further development by and for healthcare safety investigation stakeholders in England.

References

Lester, S. (2014) Professional competence standards and frameworks in the United Kingdom, *Assessment & Evaluation in Higher Education*, 39(1), 38–52

NHS England, (2022) Patient Safety Incident Response Framework, available at https://www.england.nhs.uk/long-read/patient-safety-incident-response-framework/ (accessed 25 November 2024)

Nixon, J. and Braithwaite, G.R. (2018) What do aircraft accident investigators do and what makes them good at it? Developing a competency framework for investigators using grounded theory, *Safety Science* 103 (2018) 153–161

Ensuring that UK medical graduates meet the General Medical Council's outcomes relating to Human Factors

Fraser Gold[1], Maude Adams[1], Andre J Carpio[1], Dinuki de Alwis[1], Connor Schlemmer[1], Claire Taylor[1], Paul Bowie[2] & Helen Vosper[1]

[1]University of Aberdeen, UK; [2]NHS Education for Scotland, UK

SUMMARY

The General Medical Council (GMC; the UK medical regulator) "Outcomes for Graduates" document indicates newly qualified doctors must demonstrate they can practise safely. Furthermore, they must actively participate in improvement work relating to safety and quality. The detail of this overarching outcome positions Human Factors as central to this activity. This reflects an international direction of travel that recognises the lack of success of what might be described as a patient safety 'movement': Despite collective efforts, safety has not improved. In fact, a recent run of safety 'scandals' suggests that care is increasingly precarious. In the UK, Professional, Statutory and Regulatory Bodies agree Human Factors offers the best chance for meaningful safety and quality improvement, evidenced by national initiatives such as the Academy of Medical Royal Colleges National Patient Safety Syllabus and the Patient Safety Incident Response Framework (PSIRF). Including Human Factors-related outcomes in health and care educational programmes is not a minor 'bolt on' – it requires a co-ordinated and strategic approach. It also faces several challenges, not least the lack of Human Factors competence and capacity available to most educational institutions. This case study explores how a comprehensive and sustainable Human Factors curriculum was embedded in an undergraduate MBChB (medical) curriculum in a UK university. It is believed that this is unique in the UK, and that sharing our experiences will support educational faculty in developing their own programmes. It also offers the opportunity to start a national conversation about agreed HF outcomes across health and care curricula.

KEYWORDS

Medical education, Patient safety, Human Factors competency

Introduction

The GMC Outcomes for Graduates indicate newly qualified doctors must demonstrate that they can practise safely, positioning Human Factors (HF) as central to this. This reflects an international direction of travel that recognises what it means to 'practise safely' has changed. Despite collective efforts, the patient safety 'movement' has arguably failed. A contributory factor has been the basing of safety efforts on Quality Improvement (QI), which has its roots in industrial process, more suited to manufacturing than to the complexity of modern healthcare. UK Professional, Statutory and Regulatory Bodies now agree that HF offers the best chance for meaningful improvement, evidenced by national initiatives such as the Academy of Medical Royal Colleges National Patient Safety Syllabus (and associated training of patient safety specialists) and the Patient Safety Incident Response Framework (PSIRF), both based on HF.

These initiatives, while welcome, are undermined by the lack of HF competence across the health and health education sectors. CIEHF recognised this in its 2018 White Paper[1] "Human Factors for health and social care". Underpinning the vision of HF integration was the notion that "sufficient and relevant HF education is included in clinical curricula". If HF is to be applied in practice, then "sufficient" education must go beyond knowledge to include HF competency. There is another threat to successful delivery of such education - the UK healthcare-specific conflation between HF and what might be described as 'factors of the human' (or non-technical skills). This is from a misunderstanding during the transfer of safety learning from the aviation to health sectors. Non-technical skills training was a response to a systematic analysis of a series of incidents where crew dynamics played a significant part. The 'Human Factors' part was the systems approach to recognising that an intervention addressing non-technical skills was necessary. Unfortunately, this is rarely recognised in healthcare, meaning HF has become conflated with non-technical skills. A knock-on effect is the misunderstanding that HF is a small part of QI which has contributed to the ongoing dominance of QI in UK healthcare. It could be argued that QI (in its purest sense) and HF are ideologically very different. QI seeks to strip out variation and to standardise, while HF recognises that variation results from the reality of work in a particular system. Understanding that reality is key to developing improvements that work for that context. This presents an educational challenge: we must support *learning* about HF but also *unlearning* of what went before.

Much healthcare education is beyond the control of academic staff. For safety, much of what students learn is driven by the safety-related attitudes and values they observe in the workplace (the "hidden curriculum"). This may be at odds with formal teaching, and so any educational strategy must provide opportunities for students to reflect on mismatches between the taught and hidden curricula and be accompanied by HF education and training for faculty.

Approach

Any educational strategy needs to be constructively aligned, starting at the end, with a vision of the HF knowledge, skills and competencies medical graduates should possess. Valid and reliable assessment is then designed to support this, after which learning activities can be developed to support assessment success. The strategic approach also needs to ensure that the educational programme is adequately resourced and sustainable. For this current study, a consensus development approach was taken, which facilitates sharing of expert opinion. Experts were drawn from the University of Aberdeen and from NHS Education for Scotland. Some members of the expert group have dual roles with organisations involved in the delivery of National Patient Safety Syllabus training. An early decision was to base the development of the University of Aberdeen MBChB Human Factors for patient safety curriculum on a model proposed by Vosper and Hignett[2]. Key points of this model include the observation that any curriculum exists within a complex sociotechnical system, and therefore HF should be central – not just as *content*, but also as a *methodology* underpinning the design of an accessible curriculum that optimises system performance as well as enhancing staff and learner wellbeing.

Findings

Expert consensus established the following guiding principles:

- The curriculum should support development of competencies reflecting national initiatives such as the Patient Safety Specialist training, ensuring graduates are workplace ready.
- Activity should reflect the 'competence and capacity' pyramid described in the CIEHF White Paper[1]. This suggests that effective embedding of HF needs to be overseen and supported by Suitably Qualified and Experienced Professionals; That most HF-based activity will be led by a larger group of staff who are 'relative experts'; That all staff (across

the organisation) need to understand basic HF principles to ensure their activities are not incongruent with any HF work. Consequently, curriculum design was led by chartered members of CIEHF, and two strands were developed. The Core Strand ensures all students have a high-quality HF education, while the Advanced Strand allows interested students to work towards professional recognition as a Technical Specialist (TechCIEHF).

- Those undertaking the Advanced Strand must have meaningful opportunity to apply HF in practice, requiring effective partnerships with clinicians with an HF interest.
- The curriculum should support learner engagement with the wider HF community, including CIEHF, allowing them to develop the networks that will be critical in the workplace.
- Space in the curriculum is extremely limited so a key focus should be developing existing activities, viewing them through a Human Factors lens.
- Alongside the HF curriculum, staff development must be considered to ensure sustainability.

Developing the Advanced Strand was relatively straightforward. MBChB curricula include student-selected and elective components throughout the course, and we have developed HF options for all of these. The largest is in the Year 3 Medical Humanities block, where we have developed an HF course (with 150 notional hours of student effort). Medical Humanities gives all students an alternative perspective on medicine – much of their course is science-based, taking a 'positivist' approach which in many ways is at odds with the claim that healthcare is 'person-centred.' In the Humanities block they are encouraged to understand medicine, health, sickness and disability from different perspectives including that of the patient. HF is perfectly placed in this block – it sits on the cusp between the sciences and the humanities, providing students with practical tools for understanding patient stories in a way that can lead to improvements in safety and quality. In this course, learners get extensive practical experience with HF methods/tools which can be developed in electives later in the course. The course is CIEHF-accredited, and outcomes reflect our professional competencies, helping students to map activity for their portfolio for TechCIEHF.

The Core Strand was more challenging, as it relied on opportunistic working, but this has proved surprisingly effective, and has actually benefitted from cutbacks in the sector. Reduction in relative staff numbers always comes with a redistribution of tasks, which can be an opportunity. For example, the project lead 'inherited' the Year 1 lectures on Drug Therapy. This is recognised as being a critical area for patient safety. Existing teaching focusses almost entirely on 'what the body does to the drug'. However, patients take drugs in a system that extends far beyond their physical body, and not understanding this in the round creates additional risk. In the new teaching, students view the kinetic processes of absorption, distribution, metabolism and excretion through a systems lens. A full description of this beyond the scope of this paper, but one example would be the excretion of drugs in the urine. More usually, this would be viewed largely from a kidney function perspective – if your kidneys aren't working well, then it will take longer for the drug to be cleared. If the dose isn't adjusted to account for this, then we may end up with blood levels becoming toxic. However, it is possible to end up in the same situation even when kidney function is normal. Dehydration has a significant impact on drug clearance and prescribers rarely consider patients might be deliberately dehydrating themselves. Why might they do this? If you take a systems approach and consider drug therapy from the patient perspective, it becomes apparent that deliberate dehydration is a necessary strategy for many people. For example, women who have had children often suffer stress incontinence, and deliberate dehydration may be one strategy for reducing the risk of leakage, especially when access to public toilets may be limited. Access to public toilets is an environmental factor in systems terms, influenced by even more distant external environmental factors, such as spending on public services, and perhaps vandalism (contributed to by other social factors!) Another reason for deliberate dehydration might be disability. Travel can be disabling for people with even relatively minor impairments. For example, toilet accessibility onboard aircraft is very poor, and travellers with reduced mobility often report dehydration as a strategy for coping. This approach has two very powerful messages for learners. Firstly, the need to understand the system in which the patient is taking their medication. Secondly, drug therapy is an equality, diversity and inclusion (EDI) issue. Almost invariably, those at higher risk of a medication-related adverse drug event are the poor, the old, females, the non-white and the disabled. For little extra work, we have covered the necessary drug therapy teaching, introduced the systems framework that students will return to throughout their course *and* raised EDI issues. This latter aspect is not minor: it is not only a social justice issue, but a curriculum efficiency. The GMC has mandated EDI outcomes, and this new approach to drug therapy allows these to be addressed in a meaningful way, giving us double value for the same learning activities. In the last academic session, we extended this by considering sustainability implications for deprescribing. A major contributory factor to medication-related adverse events is polypharmacy (where a patient is taking multiple medicines). Deprescribing is the planned and supervised process of dose reduction or stopping of medication that might be causing harm, or no longer be of benefit. Prescribers are reluctant (often citing safety concerns), and so it seems sensible to enhance safety by using a systems framework to support deprescribing. However, we also know pharmaceuticals are an environmental burden, from the carbon footprint associated with their manufacture through to the impact of drug contamination of aquatic ecosystems. As Barry Melia, Principal Pharmacist at Public Health Scotland (and Chair of Environmental Sustainability at the Guild of Healthcare Pharmacists) says, "the most sustainable medicine is the one that is never prescribed." Our teaching on deprescribing therefore also contributes to the sustainability curriculum.

Building capacity and capability

Alongside content development, a faculty education package was developed (also CIEHF-accredited). This forms part of the University of Aberdeen MSc Clinical Education, and is available externally as a standalone course, meaning there is the potential for revenue generation which could support further curriculum development. One possible route would be micro-credentialling, offering bite-sized chunks of HF learning as part of a continuing professional development (CPD) package.

Currently, our 'student experts' are supported in becoming HF faculty, teaching in some of the Core Strand activities. It is hoped that the early adopters will be making applications for Technical Membership in about 18 months' time. Some of their experience is captured in the case studies below.

Learner stories

Fraser Gold (Year 4)

"My first introduction to HF was through the Medical Humanities block. Before this course, I hadn't even heard of the term "Human Factors", but I had heard of ergonomists and Quality Improvement. I was interested in this course as one of my biggest fears of becoming a junior doctor is making an innocent mistake and being struck off by the GMC, making my six years of medical school worthless. I liked the idea that Human Factors considers the person in the context of a system and looks at finding more robust barriers to prevent errors rather than simply "blaming and retraining" the individual."

This is a very powerful observation. The GMC is in a difficult position: as the regulator, it is charged with holding individual practitioners to account in a way that perhaps doesn't reflect modern safety science. An attempt has been made to account for this in the GMC's Good Medical Practice document: The GMC will respond to concerns about a practitioner by considering:

- The seriousness of the concern

- Any relevant context that may impact on risk (*including systems factors*; our italics)
- How the medical professional responded to the concerns

This is good progress, but the only way those contextual factors can be understood is if the medical practitioner is able to articulate them and explain how they impacted on practice. As Fraser suggests, Human Factors is about so much more than patient safety:

“I carried out a survey on students’ mental health during the pandemic. One of my biggest worries was that if I admitted to struggling with my mental health, then the GMC may issue me a fitness to practice, and I wondered if other students felt similar. The survey revealed that 47% of respondents have worried about fitness to practice due to their mental health, and 77% have hesitated or not contacted the medical school due to this worry. Many of us do not seek support due to the fear of the GMC striking us off, and outside of medical school, healthcare workers generally do not feel protected in their workplace and worry about making mistakes. If the environment and culture of healthcare were changed to a more positive one with space for learning from mistakes using a Human Factors approach, then we may retain more staff and prevent burnout.

“This [HF] course really highlighted to me how important HF is within healthcare and how we all must be champions of it. It is mandated in the GMC ‘Outcome for Graduates’, yet many doctors remain unaware of its existence, let alone its importance. To me, patient safety is all about reducing preventable harm to our patients by strengthening barriers. A systems approach allows us to learn from incidents, which give us insight into the robustness (or otherwise) of our existing barriers. A mistake should not punish those with good intentions, who get caught up within the pressures of a system and we must protect and look after our most valuable resource within the NHS, our staff.”

Connor Schlemmer (Year 3)

“Studying HF has enhanced and shifted my views on patient safety. I began the course with a viewpoint far closer to “root cause” and have been shown the value of examining problems more holistically, understanding that system errors are the result of interactions between system factors. It altered my thinking, from more surface level to seeing things in terms of system factors, interactions and resultant wanted and unwanted outcomes. It’s made me more focused on patient safety and far more perceptive of the importance of design – not just equipment, but also systems as a whole to ensure they support the people that work in them. HF has moved me away from directing blame towards individuals, recognising that ‘fire and rehire’ doesn’t fix problems. Only a systemic analysis can result in sustainable improvement after things go wrong.”

Connor’s reflection captures one of the key intended learning outcomes of the HF curriculum: that change in mindset which is necessary to support improvement. The HF course within the Medical Humanities block requires learners to undertake a partial systems analysis of their choice. Connor’s reflections on booking a GP appointment highlight another foundational concept: Meaningful analysis of complex systems will often raise more questions than it answers:

“My project focused on the process of booking a GP appointment, following personal experiences of the difficulties. My systems analysis considered the perspectives of multiple stakeholders including doctor and patient. I identified issues in primary care such as understaffing, high workloads and lack of retention, and also considered possible recommendations and solutions. While the project identified some possible options for improving the efficiency of the primary care system at a lower level, generally it was considered that change needs to happen at a higher level, including areas such as the current NHS GP contract as well as government policy relating to the training and retention of new doctors and fully trained consultants.”

AJ Carpio (Year 3)

Most of our learners are surprised by the way a systems approach 'opens the lid' on the complexity of apparently simple tasks. Furthermore, AJ's reflection reveals another HF "truth" – it is often the mundane minutiae that undermine successful outcomes, and it is worth trying to understand these.

"I did my HF project on the pain management pathway in Aberdeen Royal Infirmary. The system consisted of many different health care professionals interacting with each other and with various system elements such as equipment, electronic databases and prescribing platforms. By analysing interactions, I identified those with a particularly strong impact on the desired outcome of rapid pain relief. Key parts of the pain management process included: the patient notifying the nursing team of pain, nursing team communicating this to the medical team and lastly the medical team prescribing medication and communicating this to the nursing team. Communication was a key theme here, but I was surprised to see how often it was affected by unexpected systems factors. For example, sometimes 'notifying nurses' meant the nurse seeing the patient was in pain. Design and layout of patients' rooms could facilitate or hinder this recognition. Windows allow staff to monitor patients from the door. However, in the emergency department, the windows were covered in opaque film. I discovered that little squares had been cut from the film (a post-installation modification) to enable observation – a perfect example of [what, in systems terms, would be called] a workaround.

"My recommendation was to improve communication at all stages of the process, and my systems analysis allowed me to suggest solutions appropriate for the context. For example, redesigning the prescribing system to automatically inform the nursing team of any changes made to a prescription would enhance communication in the final stage of the process. This potentially reduces the cognitive workload on clinicians, prevents erosion in team dynamics and also reduces the time taken for patients to receive analgesia. Having done this project, I now have a deeper understanding of patient safety. I have learned that blaming people for mistakes is not the right approach for improving patient safety, but rather redesigning the system to make it harder for people to make mistakes. When it is not possible to 'design' out errors, it is possible to include redundancies to minimise harmful events."

Claire Taylor (Year 3)

Claire's reflection builds on this idea that the seemingly unimportant can hold the key to improvement. It is the HF specialist's job to dig into the messy reality of work and consider the relative contribution of system entities. Claire's project concerned the use of Early Warning Scores (EWS) which are a mainstay of identifying critically unwell and deteriorating patients. "Failure" to identify such patients has been cited in numerous incident investigations, and the organisational attitude to such "failures" can be summed up in a recent review[3], which concludes that EWS are effective tools, so long as staff use them properly! A national investigation by the Healthcare Safety Investigation Branch (now the Health Services Safety Investigations Body; HSSIB)[4] challenged this – other aspects of the work environment made it difficult to complete the observations necessary for EWS to be effective. Claire's work adds to this:

"Manual respiratory rate (RR) measurement is vital for the early detection of patient deterioration but is often performed inaccurately in practice. This study, employing a systems approach, investigated the systemic challenges affecting RR monitoring in the NHS. The findings highlight common pressures such as heavy workloads, high nurse-to-patient ratios and time constraints, which contribute to the reliance on estimation methods (quick counting, and the preference for even numbers), practices that undermine the accuracy of RR documentation. Variability in experience and knowledge among clinical staff also contributes to inconsistent measurement practices; with those who have just recently graduated being more likely to maintain best practice. Furthermore, the study reveals that RR is often undervalued compared to other vital signs, leading to a further normalisation of inaccurate recording behaviours. This raises a critical question: if RR data is frequently inaccurate, what is the value of recording it on EWS charts? Rather than attributing these inaccuracies solely to ‘human error’ or insufficient training, the study emphasises the importance of understanding broader system influences. Unlike other vital signs, RR monitoring lacks widespread assistive technology. Could AI and digital health innovations bridge this gap? If so, their implementation must be guided by HF to ensure usability and avoid unintended outcomes. Future research should focus on the feasibility and long-term impact of integrating emerging technologies in clinical practice to improve patient safety.

“As my first HF project, this study provided an eye-opening introduction to the profound impact of human-system interactions in healthcare. It highlighted the importance of designing healthcare systems that account for the needs of the people that work in them, especially in critical areas like vital sign monitoring, where inaccuracies can significantly affect patient outcomes. This experience has fuelled my commitment to understanding and improving healthcare processes through a systems-thinking approach.”

Maude Adams and Dinuki de Alwis (Year 3)

Maude and Dinuki’s reflections are grouped together here because they capture a particularly critical aspect of Human Factors: its inherent inclusivity. Maude has identified issues relating to accessing cervical screening, a largely female issue, while Dinuki’s study revealed the technological challenges of telemedicine tended to exclude older people and those with impairments.

Maude: “As part of the HF course, this study aimed to investigate the reasons behind the low attendance at cervical cancer screening appointments in the UK, using a HF approach, aiming to offer robust recommendations. Low attendance in the programme is a growing concern due to its critical importance in early diagnosis and prevention of cervical cancer. Data was gathered from literature and qualitative interviews, and a systems framework used as to analyse the task of booking an appointment for the cervical screening programme. Several factors were identified as barriers to attending appointments. Firstly, needing to be sent a physical invitation affected individuals without a fixed address, along with younger woman who moved flats often. Secondly, needing to call and attend at specific times was made harder by individuals having busy lives and additional responsibilities. Finally, the culture around appointments, which facilitates fear and a view that if you are vaccinated or practicing safe sex then you are at a low risk of cancer and do not need to attend. This is further confounded by lack of clear information provided about the need for booking, what will happen at the appointment itself, and what will happen if the results are positive. Overall, the report concluded that the reasons for low attendance are multifactorial and cannot not be solved easily. However, by using an HF approach, I was able to suggest evidence-based changes which may make a difference. These included self-testing and online booking systems.”

Dinuki: “My introduction to HF and patient safety was in this humanities course. I had heard both terms used individually, but together, they changed my understanding. If you want to work in healthcare, you must prioritise your patients' needs, which means ensuring their safety is your top concern. However, how can you ensure that this occurs? I think that HF may contain the answer to this query. For my assignment, I chose telemedicine as it is an upcoming aspect of medicine worldwide, and I based my work on a literature review that aimed to understand the intended users and strengths and weaknesses of the currently available systems. I bounded my system by studying the pre-consultation phase, which deals with the technological requirement to have a device-led consultation remotely from the comfort of your own home. My findings suggested that technology-related shortcomings and lack of understanding and knowledge on the basic operation of electronic devices that result in people being unwilling to engage in such consultations.

I had not approached any system with the HF mindset before - so far, I have always believed that errors are only the result of personal incompetence. However, the HF approach gave me the necessary insight to understand the contribution of multiple systems factors and their interactions. Understanding this offers the potential to design a system that makes it easier for operators to do the right thing. Teaching myself to observe and think in this manner will not only help me understand situations better in the future, but help me move away from blame (whether blaming others, or self-blame)."

Key takeaways

- The University of Aberdeen MBChB includes a fully integrated HF curriculum which supports learners in achieving HF outcomes that support national workplace initiatives
- The authors believe that this is unique in the UK, and that sharing our experiences will support educational faculty in developing their own programmes.
- It also offers the opportunity to start a national conversation about agreed HF outcomes across health and care curricula.

We would like to leave the final word to Maude, who has been following the Advanced strand since she enrolled for her first year, where she completed a systems analysis of a major medication-related adverse event:

"Before I studied HF, it was easy to agree with the incorrect assumption that it was the actions of people which cause errors or accidents to occur. However, my patient safety study has significantly impacted my understanding of HF and its importance to patient safety. By applying a systems framework to real-life examples, it is very clear to see how it can directly improve patient safety. Most importantly, understanding the key principles of not blaming individuals and recognising the ineffectiveness of retraining is essential if we are able to truly learn from incidents in a way that stops them from happening again."

References

Augutis W, Flenady T, Le Lagadec D, Jefford E. How do nurses use early warning system vital signs observation charts in rural, remote and regional health care facilities: A scoping review. Aust J Rural Health. 2023; 31: 385–394. https://doi.org/10.1111/ajr.12971

CIEHF. 2018. Human Factors for health and social care. A White Paper. Available from: https://ergonomics.org.uk/resource/human-factors-in-health-and-social-care.html

Healthcare Safety Investigation Branch. 2020. Investigation report: Early warning scores to detect deterioration in COVID-19 inpatients. Available from: https://www.hssib.org.uk/patient-safety-investigations/early-warning-scores-to-detect-deterioration-in-covid-19-inpatients/investigation-report/

Vosper, H., Hignett, S. 2017. A review of Human Factors and patient safety education in pharmacy curricula: a UK undergraduate perspective with lessons for pharmacy education. American Journal of Pharmacy Education 82 (3), Article 6184. doi:10.5688/ajpe6184

A Human Factors Evaluation of the use of Patient Alerts within an Electronic Medical Records program

Camilla Rowland[1], Dr Laura Pickup[1], Fiona Spence[1] & Dr Kyle J Harrington[2]

[1]University Hospitals Bristol & Weston NHS Foundation Trust, [2]University of Nottingham

SUMMARY

A human factors evaluation of the usability, reliability including staff perception of the Accessible Information Standards (AIS) patient alerts within an acute NHS Trust electronic medical records program (EPMR). Aim of the investigation was to improve patient safety with the creation of a set of patient alert system principles to support future design of patient alerts systems within EPMRs.

KEYWORDS

Alerts, healthcare, interface design

Introduction

The Accessible Information Standards (AIS) were created in 2016 by NHS England and require health and adult social care organizations within the NHS to apply five requirements, see Table 1, to meet information and communication needs of patients with a disability, impairment or sensory loss (NHS England, 2017).

Table 1: The five requirements that NHS care and publicly funded adult social care must implement.

Requirement 1	Ask people if they have any information or communication needs and find out how to meet their needs.
Requirement 2	Record their needs clearly and in a set way.
Requirement 3	Highlight or flag the person`s file or notes so it’s clear that they have information or communication needs and how to meet those needs.
Requirement 4	Share information about people`s communication and information needs with other providers of NHS and adult social care, when they have consent or permission to do so.
Requirement 5	Take steps to ensure that people receive information which they can access and understand and receive communication support if they need it.

At University Hospitals Bristol &Weston NHS Foundation Trust (UHBW) patient alerts within the electronic patient medical records (EPMR) system are used to record and flag these patient needs. It was recognized by the Experience of Care and Inclusion Team that the digital system created challenges for staff that were not recording or making reasonable adjustment to support patient needs and this was impacting on patient safety. The aim of the research study was to use human factors methods to evaluate the current design of the patient alerts system and the objective was to produce a set of principles to inform future design of patient alerts.

Methods

Two human factors methods were chosen**.** Focus groups were held, staff were recruited from Outpatient areas across the Trust and separate groups held for clinical and non-clinical staff to gain their perception of the current design of the patient alert system. Participants were shown seven of the AIS patient alerts and asked the same set of questions, see Table 2. Focus group data was transcribed into Nvivo and a thematic analysis was completed.

Table 2: Focus Group questions.

1	**What do you understand by the meaning of this alert?** Is it meaningful? What might be the implications for the patient / what are the patient's needs?
2	**Do you feel this alert is safety critical?** How quickly does this alert need to be actioned?
3	**How well does this alert fit into your workflow?** When would be the most appropriate time to receive this alert (in relation to your task)?
4	**To what extent does this alert give you the information you need to act?** Is any information missing?

A heuristic evaluation was completed of the EPMR system by human factors experts from the Human Factors Professional in Healthcare network group. As recommended by Quiones et al (2018) a domain specific set of heuristics were created to reflect three core requirements of the EPMR, usability, safety critical and efficiency. Screen shots of the EPMR were taken and the human factors experts were asked to complete five tasks, searching for a patient, reviewing the patient's alerts, adding a patient alert, removing a patient alert and adding an interpreter requirement. Their comments were recorded along with the number of prompts they required to navigate within the interface. After each session the human factors expert was sent a link to a Microsoft Form and asked to rate each heuristic and provide additional comments.

Results

A total of twelve staff attended the focus groups. Five core themes were determined from the focus group data, with the theme of *Design* eliciting the most participant exerts (Table 3). Participants perceived the current design of patients' alert lacked clarity or information to enable a successful interaction with a patient who has information or communication needs.

Table 3: Thematic Analysis codes, sub codes and data extracts

	Code Theme	Sub-code (s)	Number of Data Extracts
1	Design	• Interface mis-design • Lack of Clarity • Prompt	99

2	Patient Requirements	• Patient Communication • Patient Need • Patient Preference	48
3	Clarity of Process	• Clear communication • Lack of Knowledge of the process	44
4	Safety Considerations	• Safety • Work Demand • Time critical	75
5	Task Dependent	• Task dependent	31

Three human factor experts participated in the heuristic evaluation. *Recognition and recall* were the heuristic which participants scored as most severe for usability; *Helpfulness* was the second most severe. Participants felt that new users would struggle to locate information and doubt their choices when moving through steps for each task.

Discussion

We believe that this research study is the first of its kind to evaluate information and communication patient alerts. Previous human factors research was sourced relating to medication patient alerts within a clinical decision support system. Some similar results were found to be applicable to the use of patient alerts within an EPMR. This research concluded with the creation of a set of six principles for the creation of a patient alerts system, see Table 4.

Table 4: Principles for the creation of a patient alert system.

1	**Write a list of system requirements, involving all users** to promote recognition and acceptance of patient alerts. Consider use of appropriate cues for all users, including temporary staff.
2	**Place patient alert information next to other relevant information,** to reduce steps required to complete the task, ensuring information is visible on all devices used.
3	**Provide help sections at all decision-making actions**, help content to reflect staff job role level of knowledge and experience
4	**Create transparent alert descriptions**, including information on the hazard, consequence and instructions to avoid the hazard.
5	**Create a clear process (system of work) for patient alerts,** including responsibilities for adding and removing alerts.
6	**Ensure patient alerts are present at the most relevant time for each job role within their workflow,** for example before a decision is made.

These principles are proposed to inform NHS health and adult social care organizations when procuring future EPMR systems. Application of these principles within an EMPR system will enable successful implementation of the Accessible Information Standards. Successful implementation of these Standards will increase patient safety, staff reliability and provide efficient use of clinical services reducing missed appointments.

References

NHS England. (2017). *Accessible Information Standards Specification v.1.1* [pdf] Leeds: Patient and Public Insight Group. https://www.england.nhs.uk/publication/accessible-information-standard-specification/

Quiones, D, Rusu, C and Rusu, V (2018) A methodology to develop usability / user experience heuristics, *Computer Standards and Interfaces,* 59, 109-129 https://www.sciencedirect.com/science/article/abs/pii/S0920548917303860

Exploring Work System Factors Contributing to Nurse Drug Administration Errors

Stacey Sadler[1], Eva-Maria Carman[2] & Selina Rizwan Ladak[1]

[1]Nottingham University Hospitals NHS Trust, [2]Trent Simulation and Clinical Skills Centre, Nottingham University Hospitals NHS Trust

SUMMARY

Double-checking medication, particularly in the paediatric and neonatal setting, is a widely used intervention for the reduction of medication administration errors and is standard practice in most UK hospitals. A systems analysis was undertaken with paediatric and neonatal nurses at a large NHS teaching hospital to understand the key challenges influencing the drug administration and the second checker process, and to support work system recommendation development. A qualitative explorative approach using the SEIPS 2.0 model was adopted for three focus groups and eight semi-structured interviews. In addition, a process map was generated and a Hierarchical Task Analysis conducted, identifying barriers and enablers specifically for the second-checker task and nurse involvement. To further validate the qualitative results, a review of reported drug administration incidences within the Family Health division was conducted. A variety of sociotechnical barriers were identified that currently hinder both the primary nurse and second nurse checker in this process. The task analysis identified differences that occur in practice when compared with the local standard operating procedure. An example of a barrier unique to the second nurse checker centred on the second nurse not challenging the primary nurse if there was a discrepancy in calculations. This work provided an enhanced understanding of the issues nurses' face and how these interact and affect their day-to day role. Recommendations arising from this work were wide ranging, highlighting that to try to resolve issues within a process, it is imperative that multiple areas are focused on.

KEYWORDS

Drug Administration, Systems Analysis, Hierarchical Task Analysis

Introduction

Double-checking medication, particularly in the paediatric and neonatal (NNU) setting, is used as an intervention to prevent or reduce medication administration errors and is standard practice in most UK hospitals. Although within the paediatric inpatient setting, systematic reviews demonstrate the high prevalence of medication errors (Gates et al., 2019, Sutherland et al., 2019), and double-checking medication is common practice. Studies evaluating the effectiveness of double checking to reduce medication administration errors have concluded there is insufficient or no evidence that the double-check process versus single checking is associated with lower rates of medication administration errors (Koyama et al., 2019, Westbrook et al., 2021).

At Nottingham University Hospitals NHS Trust (NUH), the standard is to have two registered health care practitioners, both check all steps of the medication administration process for children under 18 years (or 16 years if nursed on a non-paediatric ward). Where a calculation is involved, both practitioners must perform an independent calculation and then share their individual calculation result to confirm accuracy. Both practitioners are to sign the prescription chart. Despite this standard, there have been several, significant administration medication errors, reported within the paediatric and NNU setting, resulting in patient harm. As a result, this piece of work aimed to identify work system factors and gain a better understanding of the key challenges paediatric and NNU nurses face in their day-to-day role, which may contribute and lead to drug administration errors, to support work system recommendation development.

Method

Work system factors were explored with paediatric and NNU staff using a qualitative explorative approach to understand the key challenges and system components influencing the drug administration and second checker process. A total of three focus groups and eight semi-structured interviews were conducted with eighteen paediatric and eight neonatal nurses at NUH between October to November 2023. The Systems Engineering Initiative for Patient Safety (SEIPS) 2.0 model (Holden et al., 2013) was used to form the theoretical foundation for data capturing processes. A deductive thematic analysis (Braun & Clarke, 2006) was undertaken to determine the common themes across the focus groups and semi-structured interviews, categorising the themes into the SEIPS 2.0 model work system factors. Work system factors identified were further classified into issues that were having a direct impact on the drug administration process and those that were indirectly influencing the process.
The nursing staff covered a variety of clinical specialities within the paediatric setting. Most of the participating nurses had between two- and five-years nursing experience, although a small proportion had extensive nursing experience (twenty years and above). Additional details of the participants and sessions held are included in Table 1.

Table 1: Participant characteristics for the three focus groups and eight semi-structured interviews

Staff group	**Session**	**Number of Participants**	**Years of Registration (Range in years)**	**Median Years of Registration**
Paediatric nurses	Focus groups (n= 3)	18	2-37	4.5
NNU	Semi-structured interviews	8	<1-34	3

A review of 768 reported Datix (a web-based incident reporting and risk management software used in many hospitals in the UK) drug administration incidents within the Family Health Division were also reviewed. Possible sociotechnical factors that may have contributed to the incident were noted and a deductive thematic analysis was undertaken to determine the common themes, categorising them into the SEIPs 2.0 model work system factors to further validate the qualitative data.

In addition, a process map was generated using the qualitative data and the local standard operating procedures (SOPs) on medication administration to identify the necessary tasks to be undertaken during a two-nurse drug administration process. A Hierarchical Task Analysis (HTA) (Shepherd, 1998) was conducted to create a detailed representation of each action required in each step of the drug administration process. The HTA was then used to map the identified barriers and enablers specifically for the second-checker task and nurse involvement.

Results

A wide variety of sociotechnical barriers were identified that currently hinder both the primary nurse and second nurse checker in the drug administration process. These included nurse fatigue (SEIPS: person component), managing the demands and interruptions from patients, their relatives and other health care professionals (SEIPS: person component) and a lack of NUH hand-held devices (SEIPS: Tools and Technology component). Poor internal environment factors included high temperature, noise, overcrowding and excess medication stocked within the clean utility rooms and nursing bay (SEIPS: Internal Environment component), along with a lack of nursing staff and managerial support (SEIPS: Organisation of Work component) created a challenging environment for nurses. Additional information around the identified sociotechnical barriers can be found in Table 2.

The findings on the barriers identified in the focus groups and interviews were supported by the results from the analysis of the 768 Datix reviews. Where incidents identified work system barriers as contributing to medication errors, these mirrored the sociotechnical factors identified in the qualitative data. The most common sociotechnical issue cited attributing to an error was sub-optimal staff levels (SEIPS: Organisation of Work component). Poor communication between staff and patients/relatives (SEIPS: Person component), poor nursing handovers (SEIPS: Task component), along with high and competing workload (SEIPS: Task component) and staff inexperience (SEIPS: Person component) were also frequently listed as potential contributory factors to a drug administration error.

Table 2. Sociotechnical Barriers Identified in the Drug Administration and Second Checker Process

Work system	Barriers
Person – Health Professional	Interruptions, staff fatigue, staff confidence
	Poor handwriting on medication cards
Person - Patient	Acuity of patient, Patient not wearing wrist bands
Person - Family	Pressure and interruptions from family
Task	High and competing workload and demands
Tools & Technology	Lack of devices, difficulty locating necessary equipment e.g. reference sources & keys
	Difficulty using and interpreting digital app
	Medication cards – off ward, multiple cards in use
Internal Environment	Utility room location and layout – crowding, not fit for purpose
	Ward layout and environment – crowding, high noise levels & temperature
Organisation	Staff levels, loss of experienced staff and skill mix of staff
	Poor communication, assignment of staff tasks
	Pharmacy opening times at weekends and out of hours, clinical support out of hours
	Timing of organisational procedures e.g. Delivery of intravenous Total Parenteral Nutrition (TPN), nurse shift patterns
	Cultural attitude to questioning senior staff
	Lack of support from senior management, expectations of senior staff
External Environment	Time of time e.g. day vs night

Conversely, nurses identified enablers within the work system. These included patients being well known to the nursing staff e.g. a long-stay patient or a frequent attender, as nurses felt drug administration was easier due to familiarity with the patient's medication regimen. Parents supporting the nurse by undertaking oral medication administration to their child where applicable or parents understanding the ward environment and making less demands and interruptions on the nurse also helped to free up a nurse's time. A suitable skill mix of nurses (including agency staff) on a shift allowed all nurses to undertake second checks and administer intravenous medications in a timely manner. A nurse's confidence to challenge the first nurse if they felt mistakes had occurred was also viewed as an enabler to the second check process.

The HTA identified differences that occur in practice when compared with the local SOP. These included where medication was prepared (SEIPS: Internal Environment component), preparing medication before the second nurse check occurred, the second nurse not always witnessing medication administration and not signing the medication card immediately post administration (SEIPS: Task components). Interviewed nurses knew the correct two–person check procedure but consciously deviated from the recommended procedure to enable them to fulfil other tasks. This highlights the everyday 'trade-offs' that nurses must make to enable them to complete all their patient related tasks. The trade-offs staff made should not be seen as deviations from the SOP as a choice, but rather because of trying to balance the demands of the work system, the task requirements and potential risks. As a result, the possibility of trade-offs should be considered when designing SOPS to enhance safety.

Key issues identified directly impacting the drug administration process centred around the lack of nursing staff, staff mix on a shift (SEIPS: Organisation of work component), competing nursing staff tasks and multiple high acuity patients on the general wards (SEIPS: Task and Organisation of work components). Other direct issues included suboptimal medication preparation area (either the clean utility room or preparation by the baby's bedside), which were often too hot and crowded, conditions on the ward (SEIPS: Internal Environment component), poorly written medication cards and interruptions by other healthcare professionals, family and patients (SEIPS: Person component).

Issues identified by nurses which indirectly impacted the drug administration process included lack of management support, lack of training for staff returning after taking maternity leave/career breaks (SEIPS: Organisation of work component) and not addressing potential language barriers for nurses where English was not their first language (SEIPS: person component). Inadequate nursing breaks were commonplace (SEIPS: Organisation of Work component), with nurses often reporting there was not a suitable staff room, which increased nurses' fatigue and mental stress.

Barriers unique to the second nurse checker centred on the second nurse not challenging the primary nurse if there was a discrepancy in calculations, often due to confidence issues with the second nurse checker or, cultural differences where a junior second nurse checker would not challenge a senior colleague (SEIPS: Person component). The nurse second checker provides a safety check, if the above occurs then a safety component of the drug administration process has been breached. However, the second nurse check was perceived as a positive safety measure in the drug administration process, and it was recognised as an independent check.

A range of recommendations were made to address the different work system barriers identified. These included internal work environment considerations and included reviewing stock lists on each paediatric ward and neonatal areas to help try to optimise medication storage and aid selection of medication for nursing staff, along with temperature audits on clean utility rooms, so areas that were too hot could be targeted. Tools and technology recommendations included a range of interventions from increasing the number of handheld devices and chargers on Paediatric/Neonatal wards to the introduction of Electronic Prescribing and Medicines

Administration (EPMA) within paediatrics to circumvent issues identified around handwritten/unclear medication cards and the availability of medication cards on wards. Organisational considerations included increased training for new and returning staff, setting up a paediatric/neonatal working group with the aim of further exploring which medications could be made exempt from a second check and/ or two nurse administration with the long term objective of trialling a paediatric ward with any outcomes, and a review of patient acuity versus staff level ratios to determine the optimal staff level mix on each ward. Finally, to address the person component barriers identified, a recommendation to introduce cultural awareness courses for all nurses to try and increase understanding and confidence around challenging more senior nursing staff if second checker was made.

Discussion and Conclusion

This work aimed to gain an understanding of how errors occur from a systems perspective when undertaking medication administration requiring a two-person check, rather than just attributing blame onto individual staff. The SIEPS 2.0 model, provided a structure for exploring the role of wider system factors, gaining an improved understanding of the issues that nurses' face and how these interact and affect their day-to day role. The benefit of using an approach that moves away from individual staff blame and focuses on making improvements to the work system has been recognised, in that the new NHS England Patient Safety Incident Response Framework (PSIRF; NHS England, 2024) now adopts this model. It is acknowledged, that the themes identified are from a small number of participants and as such may not be representative of the complete population for this staff group, although every effort was made to get representatives from all the paediatric specialities, and this was largely achieved. A review of Datix incidences also identified similar work systems factors to those identified in the qualitative data, adding further robustness.

The work has also highlighted the implications of nurses' high workload on the drug administration and second check process. Nurses interviewed knew the correct two–person check procedure but did deviate from the recommended procedure to enable them to fulfil other tasks and highlights the everyday 'trade-offs' that nurses must decide to make, to enable them to complete all their patient related tasks. This is reflective of the Efficiency – Thoroughness Trade -Off (ETTO) Principle (Hollnagel, 2009), where nurses are consciously balancing thoroughness (following all checks) and efficiency (ensuring timely care). Whilst this is a conscious decision taken by nurses, a lack of time and high workload means they must apply their own risk assessments in day-to day tasks. These trade-offs staff make should not be seen as deviations from the SOP as a choice, but because of trying to balance the demands of the work system, the task requirements and potential risks. This further highlights the need to understand the wider system factors that influence this task.

A variety of sociotechnical barriers were identified that currently hinder both the primary nurse and second nurse checker in the drug administration process, with wide ranging recommendations arising from this work involving different work-based systems. This highlights that to try to resolve issues within a process, it is imperative that multiple areas are focused on, rather than just one area and that a systems approach is adopted. Interventions that solely focus on behaviour change will not be successful, as the causes for workarounds will not have been addressed. This is supported by the hierarchy of effectiveness model which illustrates which risk-mitigation strategies are more effective than others in addressing factors, with people focused interventions less effective than system-based interventions (McDaniel, 2024).

It was acknowledged that some of the recommendations would be more challenging and take longer to achieve and involve different divisions within the hospital, but a variety of work system interventions have started to be implemented. These include setting up a working group of nurses and other healthcare professionals to look at the second check process. The second nurse check is still perceived as a positive safety measure in the drug administration process and will remain, but work remains ongoing around defining best practice in all clinical areas of the hospital and reviewing the current medication that requires a second check with the aim to reduce the number of medications that require second checks. Stock levels on all paediatric wards have been reviewed and rationalised. This has reduced excess stock, reduced waste and improved the working environment within the clean utility rooms, enabling nursing staff to locate medication more easily. A recruitment drive has improved nursing staff numbers within the paediatric setting and an increased amount of training and support from educational development nurses for ward-based staff has occurred, with the focus on reassuring junior nurses that it is acceptable to challenge and question calculations. It is imperative going forward that any improvements made, have nursing staff input to improve the likelihood of success and improve buy in. Regular reviews of the work system to identify any possible effects of the interventions (both positive and negative) will be required.

References

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.

Gates, P. J., Baysari, M. T., Gazarian, M., Raban, M. Z., Meyerson, S. & Westbrook, J. (2019). Prevalence of medication errors among paediatric inpatients: systematic review and meta-analysis. *Drug Safety*, 42,13–25.

Holden, R. J., Carayon, P., Gurses, A. P., Hoonakker, P., Schoofs Hundt, A., Ozok, A., & Rivera-Rodriguez, A. J. (2013). SEIPS 2.0: a human factor s framework for studying and improving the work of healthcare professionals and patients. *Ergonomics*, *56*(11), 1669–1686.

Hollnagel, E. (2009). The ETTO Principle: Efficiency-Thoroughness Trade Off: Why things that go right sometimes go wrong. 1st ed. Aldershot UK: Ashgate.

Koyama, A. K., Maddox C-S. S., Li, L., Bucknall, T., & Westbrook, J. (2019). Effectiveness of double checking to reduce medication administration errors: a systematic review. *BMJ Quality and Safety,* 1-9

McDaniel, C. (2024). The Hierarchy of Intervention Effectiveness [online]. Patientsafe – Implementing Effective Safety Solutions. Available at: https://patientsafe.wordpress.com/the-hierarchy-of-intervention-effectiveness/ [Accessed 13 February 2025]

NHS England– Patient Safety Incident Response Framework. Published 23.7.2024. Available at: NHS England » Patient Safety Incident Response Framework [Accessed 13 February 2025]

Shepherd, A. (1998). HTA as a Framework for Task Analysis. *Ergonomics*, 41 (11), 1537-1552.

Sutherland, A., Phipps, D. L., Tomlin, S., & Ashcroft, D. M. (2019). Mapping the prevalence and nature of drug related problems among hospitalised children in the United Kingdom: a systematic review. *BMC Paediatrics*. 19(1), 486.

Westbrook, J. I., Ling, L., Raban, M. Z., Woods, A., Koyama, A. K., Baysari, M. T., Day, R. O., McCullagh, C., Prgomet, M., Mumford, V., Dalla-Pozza, L., Gazarian, M., Gates, P. J., Lichtner, V., Barclay, P., Gardo, A., Wiggins, M., & White. L. (2021) Associations between double-checking and medication administration errors: a direct observational study of paediatric inpatients. *BMJ Quality and Safety.* 30(4), 320-330.

FRAM: A boundary object to understand management of paediatric leukaemia patients

Nicholas Seaton[1], Julie Crawford[1], John Moppett[2], Laura Pickup[1]

[1]Patient Safety Team, University Hospitals Bristol & Weston NHS Foundation Trust, [2]Paediatric BMT, Haematology & Oncology Department, University Hospitals Bristol & Weston NHS Foundation Trust

---

SUMMARY

This paper describes the use of the Functional Resonance Analysis Method (FRAM) to understand the potential for variability in the delivery of chemotherapy across distributed care providers. The complexity of the system is considered and how the use of a FRAM model enabled cross disciplinary collaboration. This enabled consideration to common scenarios and incidents, where variability in the delivery of care is critical to adapt to a patient's condition. The use of scenarios and incident-based analysis revealed how core functions within the system influenced the ability of teams and families to effectively monitor and communicate treatment regimes. In using the FRAM model to facilitate discussions, the team were able to challenge their beliefs on how knowledge of treatment regimens and a shared understanding across geographically distributed sites was achieved. This created new and shared knowledge on the potential variability in communication and transfer of information, essential to the reliability of the delivery of chemotherapy as intended. The paper considers the value and role of the FRAM model as a boundary object, an artefact that can support staff with different roles within the team to engage equally to solve a problem or safety concern.

KEYWORDS

Healthcare, paediatrics

Introduction

Children with leukaemia who live outside regions offering specialist cancer services benefit from a shared care system involving multiple healthcare providers, allowing for more accessible care closer to home while drawing on the expertise of regional tertiary care. However, this model of distributed care increases the complexity of the system. There is a dependency upon families to become highly knowledgeable of drug combinations and information to be accurately delivered to enable distributed cognition across clinical professionals. The risks associated with managing chemotherapy, include potential medication incidents such as missed or excessive doses, and complicated or miscommunication of care adjustments between multiple providers and families.

This paper describes work to analyse potential variability using the Functional Resonance Analysis Method (FRAM) (Hollnagel, 2012) to consider various patient safety scenarios and incidents in the delivery of chemotherapy. The work described in this paper was completed to identify opportunities to enhance resilience of the core functions relied upon for the reliability of chemotherapy treatment provided across different providers. This was achieved through team-based and organisational learning undertaken at a large acute children's hospital.

Method

The Functional Resonance Analysis Method (FRAM) was used to examine chemotherapy provision in a paediatric haematology and oncology service in the South West region of England. The investigation plan involved a qualitative data collection approach (McGill et al., 2023). Information was gathered through a document review of organisational data, such as standard operating procedures and clinical guidelines, to extract initial functions and characterise aspects. As this was used to inform a patient safety incident investigation, an ethnographic approach was adopted, involving observations of clinical processes to understand multidisciplinary team (MDT) meetings as well as the work environment to gain a deeper understanding of context and work-as-done. Semi-structured interviews were conducted with 11 multidisciplinary team members, covering 660 minutes, to understand how everyday work is performed. Parents were also interviewed for a total of 270 minutes which assisted in validating information provided by the clinical team members, to extract further functions and elaborate on aspects involving families. The interview questions focused on how chemotherapy regimens were communicated, recorded, and adjusted, as well as challenges in maintaining continuity of care.

Qualitative data analysis was thematically conducted using an inductive approach, drawing from interviews, observations, and organisational data (Braun & Clarke, 2022, pp. 55–56). This primarily followed a semantic analysis, as coding was based on explicit descriptions of work-as-done; however, some latent elements were incorporated, drawing on contextual understanding to extract and define the aspects within each FRAM function. The analysis informed how intended variabilities were essential to adapt to the patient's presentation and where less intended variability may occur. Functions were extracted from the data, and a FRAM model was developed to explore the couplings, dependencies, and variability across these functions. Two workshops, totalling five hours, were conducted with key stakeholders, including clinical representatives, to review and refine the model.

The FRAM Model Visualiser was used to design the model. Functions were colour-coded as follows: blue for Principal Treatment Centre functions, green for handover to parent or shared care unit functions, yellow for parent/caregiver functions, and orange for new functions added during the workshop.

Results

Preliminary findings revealed the complexity in paediatric care, involving multiple stakeholders across different locations, with technology playing a key role in communication and support for distributed cognition to ensure safety of the delivery of care. The electronic system for prescribing and administering chemotherapy was inadequate to reliably alert staff to incorrect or incomplete actions during safety-critical tasks. Parents and guardians were recognised as providing the resilience in the system, as they adapted to gaps in knowledge across teams exacerbated by a lack of interoperable electronic patient record systems and shared critical information between geographically distributed units. They were required to understand their child's chemotherapy regimen in considerable detail to collect and reconcile medications, but the lack of a user-friendly treatment schedule hindered their understanding and the coordination between families, staff, and care sites. Allocation of staff rosters had the potential to disrupt care continuity, with a trade-off required between extended shift duration and increased frequency to achieve continuity. This highlighted the absence of organisational fatigue management to support the reliability and effectiveness of the delivery of chemotherapy treatment. The review of the FRAM-built model also highlighted to the team a high dependency on one particular critical function (e.g., allocating chemotherapy at the Principal Treatment Centre prior to prescribing chemotherapy) not fully recognised prior to the workshops. This informed the team on the improvements they agreed to test, which were intended to enhance the ease, reliability and visibility in modifications of treatment regimes between clinicians and across clinical sites.

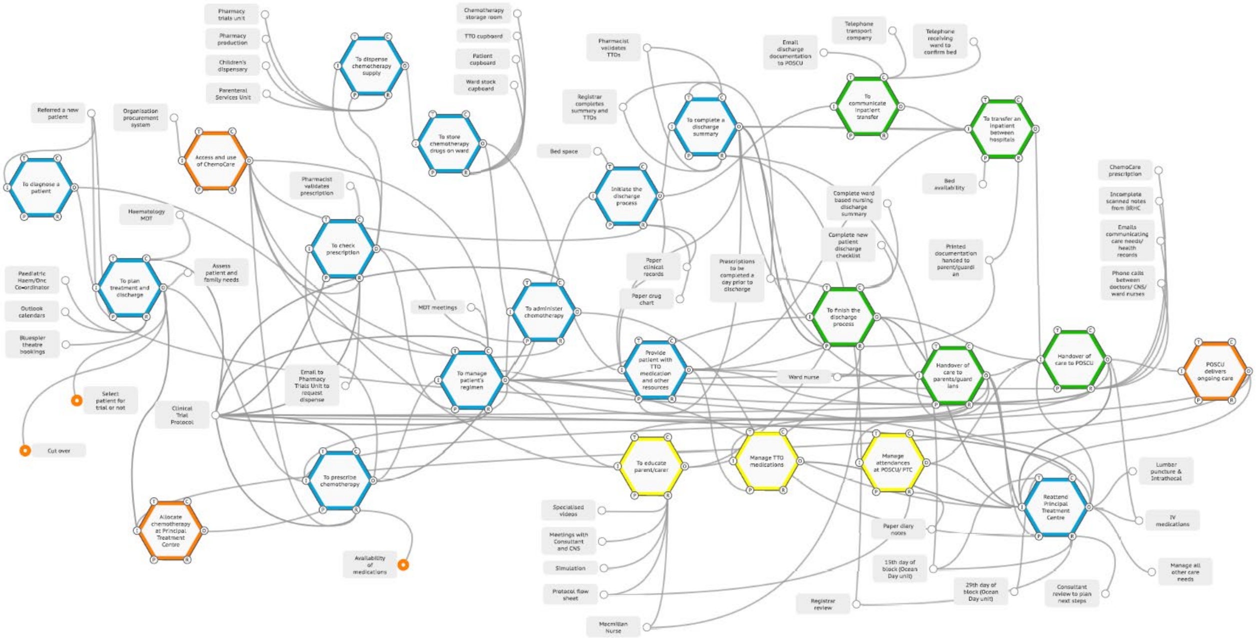


Figure 1: FRAM model of treatment regimes

Discussion

FRAM effectively visualised the complexity in the need for the system to accommodate adjustments to treatment regimes and ensure distributed cognition across the chemotherapy care system. The approach highlighted key functions and constraints shaped by geographical and technical factors. It raised important questions about safety oversight, in managing dose adjustments across locations and improving visibility in a distributed care network.

Traditional techniques used in healthcare are designed to understand how control has been lacking in the system when things go wrong (e.g., Larouzee & Le Coze, 2020). However, FRAM seeks to understand conditions of performance variability and how resilience can be built into the system (e.g., Sujan, 2021; Sujan et al., 2023). FRAM offered a 'blank model' for clinicians to visualise a systems view of clinical scenarios and incidents. This approach facilitated validation of the model and identified additional functions, which contributed to the variability in the delivery of care and required focus to enhance system resilience to accommodate these known variabilities. The process benefitted from a highly engaged clinical team and informed safety recommendations to enhance resilience locally and regionally.

The authors propose that the FRAM-built model acted as a boundary object, defined as '… a shareable and tangible artefact around which group members can interact about a problem situation of concern…' (Star and Griesemer, 1989). The workshop provided new insights to how and why the system usually succeeds but why unintended outcomes occur. The presentation of system complexity provided a shared framework that aided meaning-making between system investigators and clinical staff (Nathues et al, 2024). Using models as boundary objects is recognised as supporting the development of new knowledge through a collaborative approach to problem solving (Franco, 2019). The FRAM model from this work will be used to openly explore incidents and proactively simulate future scenarios and system changes, supporting further interventions. The limitations of this work and future consideration will be given to how families can contribute to system reviews.

References

Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. SAGE.

Franco, L.A. (2013). Rethinking Soft OR interventions: Models as boundary objects. European *Journal of Operational Research, 231*(3), 720–733. https://doi.org/10.1016/j.ejor.2013.06.033

Hollnagel, E. (2012). *FRAM: The Functional Resonance Analysis Method: Modelling Complex Socio-technical Systems* (1st ed.). CRC Press. https://doi.org/10.1201/9781315255071

Larouzee, J., & Le Coze, J.-C. (2020). Good and bad reasons: The Swiss cheese model and its critics. *Safety Science*, *126*, 104660-. https://doi.org/10.1016/j.ssci.2020.104660

McGill, A., McCloskey, R., Smith, D., Salehi, V., & Veitch, B. (2023). Building a Functional Resonance Analysis Method Model: Practical Guidance on Qualitative Data Collection and Analysis. *International Journal of Qualitative Methods, 22*. https://doi.org/10.1177/16094069231211145

Nathues, E., van Vuuren, M., Endedijk, M. D., & Wenzel, M. (2024). Shape-shifting: How boundary objects affect meaning-making across visual, verbal, and embodied modes. *Human Relations, 0*(0). https://doi.org/10.1177/00187267241236111

Star, S. L., & Griesemer, R., J. 1989. Institutional Ecology, 'Translations', and Boundary Objects: amateurs and professionals in Berkeley's Museum of Vertebrae Zoology. *Social Studies of Science, 19:* 387-420. https://doi.org/10.1177/030631289019003001

Sujan, M. A. (2021). Muddling through in the intensive care unit – A FRAM analysis of intravenous infusion management. In J. Braithwaite, E. Hollnagel, & R. Wears (Eds.), *Resilient health care, Volume 6* (pp. 101–106). CRC Press. https://www.routledge.com/Resilient-Health-Care-Volume-6/Sujan-Braithwaite-Wears/p/book/9780367338407

Sujan, M., Pickup, L., de Vos, M. S., Patriarca, R., Konwinski, L., Ross, A., & McCulloch, P. (2023). Operationalising FRAM in Healthcare: A critical reflection on practice. *Safety Science*, *158*, 105994. https://doi.org/10.1016/j.ssci.2022.105994

Introducing CoolSticks for anaesthesia; a human factors approach

Joseph Swani, Paul Southall, Frances Ives, Shakira Nathoo & Rachael Cresswell

Worcestershire Acute Hospitals NHS Trust

SUMMARY

Regional anaesthesia is commonly tested using ethyl chloride spray, but it is harmful to the environment. The CoolStick is a cost-effective alternative with a lower carbon footprint. This project used human factors methodology in the implementation of CoolSticks within a hospital, aiming to achieve a safe and effective transition and to reduce ethyl chloride use.

KEYWORDS

CoolSticks, Anaesthesia, Healthcare

Introduction

When a patient has a regional anaesthetic (for example a spinal or epidural anaesthetic), the aim is to numb a body area, either for pain-relief or for surgery. To check if the anaesthetic is working, it is commonly tested whether the patient can feel cold and touch sensation on their skin. Ethyl chloride spray is a vapo-coolant presented in a disposable cannister, it normally feels cold when it is sprayed on the skin and is one of the most common methods used in the United Kingdom for testing regional anaesthesia. However, ethyl chloride is considerably harmful to the environment, animals and plants. The CoolStick is a cost-effective alternative with a lower carbon footprint and less waste. It consists of a stainless-steel body and a plastic handle, it's reusable and stored in the fridge. This project aimed to transition to using CoolSticks instead of ethyl chloride in a hospital's operating theatres. Given that this would be a significant change in practice, we used human factors methodology with the aim to achieve a safe and effective transition.

Methodology

The project was initiated by one of the authors, who is a consultant anaesthetist with a specialist interest in sustainability in healthcare and an Environmental Advisor to the Royal College of Anaesthetists. Funding was applied for and granted from Greener National Health Service (Healthier Futures Fund). A working group was created which included an anaesthetic doctor training in human factors and a Chartered Human Factors Specialist experienced in using human factors to support innovation within healthcare. The group formulated a plan for how human factors would be used to support the implementation of CoolSticks.

Hierarchical Task Analysis (HTA)

HTA was used to compare the tasks and sub-tasks involved when using a CoolStick versus using ethyl chloride. Using HTA allowed us to recognise the increased task complexity in gathering and putting away equipment when using a CoolStick in comparison to ethyl chloride. These can be seen below (Figure 1 and Figure 2).

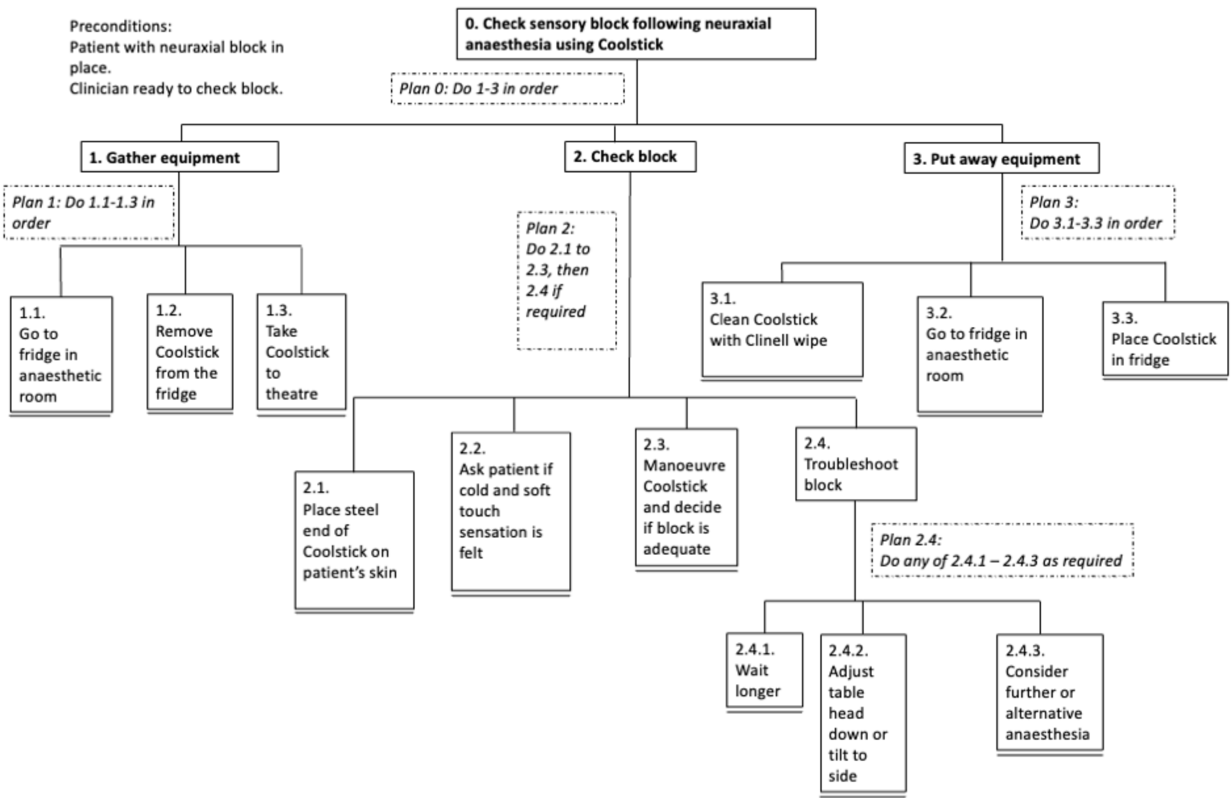


Figure 1: HTA using a CoolStick

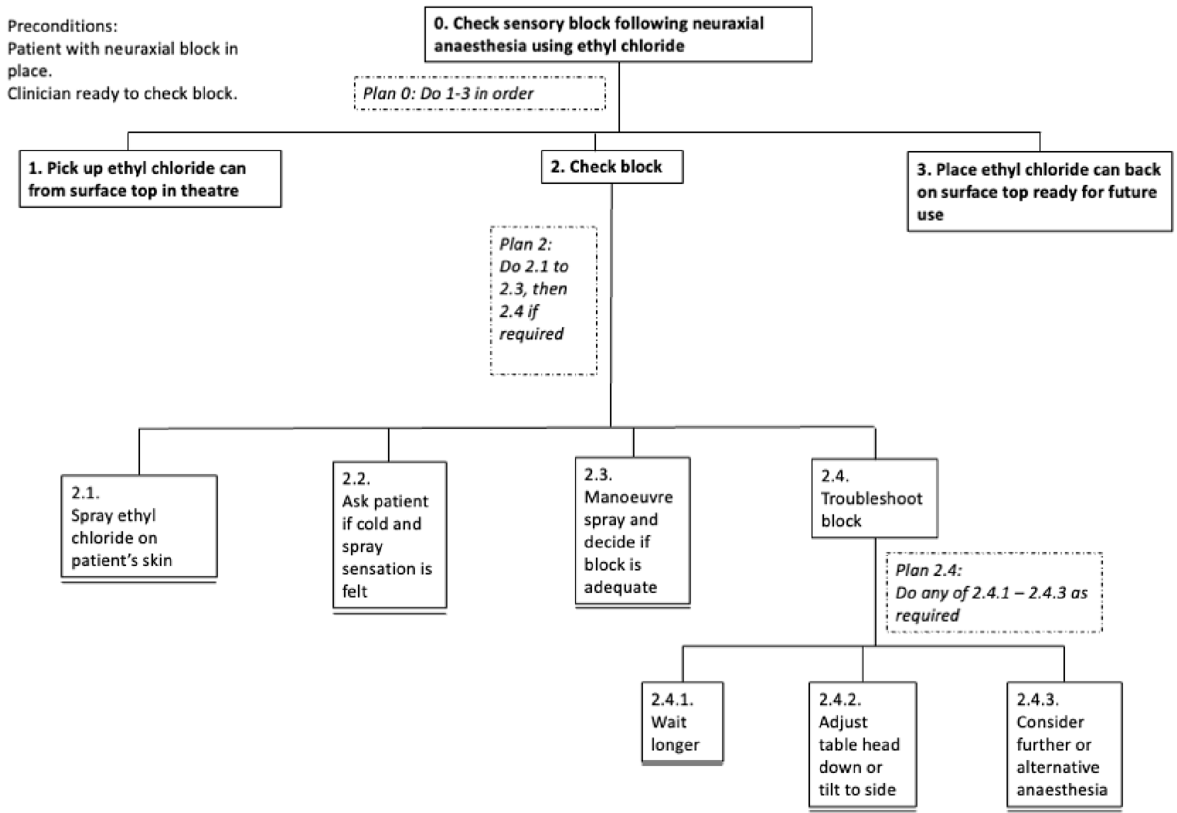


Figure 2: HTA using ethyl chloride spray

Failure Mode and Effects Analysis (FMEA)

Using FMEA, the steps in the process (defined by the HTA) were analysed to predict where and how the implementation of CoolSticks may fail. Consequences were predicted to have the potential to include using ethyl chloride spray instead, contamination of equipment and incorrect analysis of adequacy of anaesthesia. Actions were advised to try and prevent such failures.

Staff Training Video

A training video was created, demonstrating how to use the CoolStick. A video was used to allow staff to view the information easily and in a short space of time (<10 min length). It was shared using a quick-response (QR) code displayed on posters distributed throughout the anaesthetic department and theatres, in locations where CoolSticks are used. The video is continuously available so staff do not have to rely on memory and so new staff can view it without the need or reliance on training sessions. Screenshots from the video can be seen below (Figures 3, 4 and 5).

Figure 3: A CoolStick

Figure 4: Testing sensation using a CoolStick

Figure 5: Cleaning the CoolStick

Observation of practice

Following the introduction of CoolSticks into practice, time was spent in obstetric theatres to observe anaesthetic teams and their testing of regional anaesthesia. Where possible, observation was done without informing the team and without interruption to try and reduce observer influence on staff behaviour and minimise the gap between work-as-observed and work-as-done. Observations were done in the daytime and overnight including situations of emergency anaesthesia to try and capture all aspects of work. Findings provided valuable information on CoolStick and ethyl chloride use, including how staff may make adaptations to tasks. Ethyl chloride spray was found to often still be on the anaesthetic machine and sometimes used rather than a CoolStick. On occasion staff would even pick up the ethyl chloride spray through habit, then put it down (without using it) after remembering about CoolSticks. CoolSticks were most commonly removed from the fridge at the immediate time of need, but not placed back in the fridge immediately after use. Instead, after using the CoolStick it was more commonly placed on the surface of the anaesthetic machine rather than being put back in the fridge. The most common time to return the CoolStick to the fridge was at the end of the case, after the patient had left theatre. This adaptation is perhaps done to minimise the need to leave the theatre room to go to the anaesthetic room at an important time of anaesthesia, where there are multiple considerations for the patient. It allows anaesthetists to stay in the theatre room, the trade-off is that the CoolStick will remain out of the fridge for longer and become warmer. Additional observations included a variation in how the CoolStick was placed on patient skin for testing, mostly being placed in discrete areas but some would roll or slide the CoolStick along the skin in one motion.

Staff interviews

Anonymous discussions were conducted with involvement of multiple members of the multidisciplinary team including consultant anaesthetists, trainee anaesthetists and operating department practitioners. Information was gathered on perceived usability comparing CoolSticks to ethyl chloride spray. There were felt to be advantages and disadvantages to each, but overall those who had been using CoolSticks regularly found them to work well. As ethyl chloride spray has been used for many years, staff recognised that this change in equipment is not an easy transition (changing their routine and long-term practice) and this can affect their initial feelings about the new equipment. System challenges were explored including the importance of equipment location. It became clear that the location of ethyl chloride spray in the theatre room is a significant reason for some choosing to use it instead of CoolSticks, in fact the visibility of ethyl chloride spray in theatre can feel like a prompt to use it. Having to leave the theatre room and go to the anaesthetic room to get and put away CoolSticks is a barrier to its use. As not returning the CoolStick to the fridge for a prolonged period of time could mean it warms above the temperature required for cold sensation testing, having multiple CoolSticks is important so that they can be rotated from the fridge. When doing this, some use workarounds to ensure they do not re-select the previous CoolStick. These include not putting a CoolStick back in the fridge until a new one is taken out for the next case or returning a CoolStick to the bottom of the pile and selecting the next CoolStick from the top of the pile.

Learning points

Using human factors methodology in our project was tremendously beneficial by facilitating a focus on system evaluation and adaptation rather than individual staff training. What could be predicted as a simple introduction of basic equipment was found to be much more complex. Involvement of staff in the process rather than forcing change upon them was very important for staff engagement. Observations and staff interviews allowed us to get closer to understanding work-as-done. The main recommendations from the project are focussed on the design of the environment, where having CoolSticks located in a fridge in the operating theatre rather than the anaesthetic room would mitigate a lot of the barriers to the gathering and putting away of the equipment, which are the aspects with more complexity in comparison to ethyl chloride. Additional recommendations included the availability and location of cleaning wipes, the storage location of ethyl chloride as an alternative and updates to the training video based on feedback. There has been a 75% reduction in ethyl chloride use in main theatres, which will have significant environmental benefit. In addition, there is financial benefit with a projected saving of around £13,000 per annum. There is ongoing work to try and improve CoolStick use in other areas of the hospital. The learning from this project is very relevant for other hospitals, as they are likely to implement CoolSticks in the future once the financial and environmental benefit is realised.

Local rationality question tool: understanding why it made sense at the time

Louise Roe

Maternity and Newborn Safety Investigations (MNSI) programme

SUMMARY

Local rationality describes how people make decisions, based on what made sense to them at the time (Eurocontrol, 2014). While the importance of understanding local rationality in safety investigations is acknowledged, there is little to support safety investigators with how to effectively gain this knowledge when meeting with staff involved. A local rationality question tool was developed to help healthcare safety investigators understand why an action or inaction made sense at the time, without staff feeling blamed or interrogated.

KEYWORDS

Local rationality, decision making, tools and techniques

Introduction

In the field of safety investigation, it is vital to explore local rationality and fully understand why a decision, action, or inaction made sense at the time, without making the staff involved feel interrogated or blamed. In practice this can be easier said than done, and a trade-off can occur; where either the quality of information gained is sacrificed in fear of interrogating staff, or their psychological safety is neglected to get the information needed for the investigation.

In healthcare, the impact of a patient safety incident is devastating for patients and their families. Alongside this, the emotional effects on the staff involved are well documented (Wu and Steckelberg, 2012). Due to the nature of the healthcare industry, it is observed that staff will regularly individualise blame on themselves following a safety incident. The role of the healthcare safety investigator is not to add to this burden further by the way that questions are asked.

Understanding a situation from the perspective of those involved, to include their mindset, knowledge, demands, goals, the context of a situation, and any other information available to them at the time, is the essence of local rationality (Eurocontrol, 2014; Dekker, 2014). The safety investigator needs to be mindful of this and not use hindsight to judge what they think should have been obvious at the time, because it may well not have been.

The Maternity and Newborn Safety Investigations (MNSI) programme is part of a national strategy to improve maternity safety across the NHS in England. When undertaking maternity healthcare safety investigations, it was identified that whilst there was a clear understanding of the importance of local rationality, there was little information to support healthcare safety investigators to ask the *right* questions, in the *right* way, for both the investigation and the staff involved. In response to this a local rationality question (LRQ) tool was developed.

Method and approach

A literature review was undertaken and 8 English language journal articles and textbooks by safety science experts were selected. Literature selection was based on content that included decision-making, cognitive ergonomics, and incident investigation. Existing questions from safety science experts were selected and underpinning theories on factors that impact decision-making and task performance were used to provide supporting information. Questions were chosen on their ability to facilitate exploring 'why it made sense at the time' while considering the emotional impact on the staff involved and to prevent them from feeling interrogated or judged. Leading or influencing questions were avoided by adapting them, where necessary, using the tell, explain, describe approach. Some phrases were altered to ensure they were not accusatory. The tool's purpose was to bring together a collection of questions, in one place, that healthcare investigators could use to explore local rationality effectively and easily when speaking with staff involved in a patient safety incident.

The final selection of questions was collated into a functional table design (table 1) and categorised into areas of focus, to allow easy navigation in real-time practice. Categories included situation, thoughts and decision-making, preparedness, communication, and anticipation/thinking ahead. Supporting information was added to the table to make it easier to select the right questions for the information needed. The use of the tool was trialled during the author's investigations, and improvement in the quality of information obtained was evident, along with increased confidence and positive experience when speaking with staff. The selected questions were later peer-reviewed to ensure they met the needs of MNSI investigators. The tool was implemented for optional use during MNSI investigations, before being shared widely with healthcare stakeholders and safety investigators from other industries.

Table 1: Local rationality question tool

No.	Question	Comments
	Situation	
1	Describe to me what was happening at the time?	Allows exploration of dynamic elements of a work environment.
2	Describe to me what was happening around you?	How did the situation unfold around them; what cues did they get when?
3	If you had to describe the situation to your colleague at that point, what would you have told?	
4	Describe to me what you were seeing/hearing?	
5	Tell me what the workload was like for yourself and those around you during this time?	Workload intensity or inactivity. Helps to explore fatigue and stress as well. Were their multiple goals at the same time?
6	Can you tell me about any time pressures to complete tasks or any limitations on what you were able to do at that time?	Workload intensity.
7	Tell me about any reassessments of the situation?	Immediate feedback, careful monitoring and assessment of the situation. Shifting goals as patient's condition changes and new problems/complications arise.
8	Can you breakdown … (name the task) into three to six steps. Of these steps, tell me a bit about those that required assessment/decision making/problem solving, if any.	Complete task diagram for very complex procedures/situations. This can provide road map and inform subsequent

		interviews. Helps to explore situations with dynamic quality to them.
	Thoughts/decision making	
9	Describe what you were focusing on?	
10	Can you explain what your aim (goal) was during this time?	Not knowing the outcome, you now know about now. Selecting information to confirm their current hypothesis rather than explore others.
11	Tell me what was the aim when undertaking …	What were they trying to achieve.
12	Can you tell me about any previous experiences you have had in similar situations?	Pattern matching – always previously turned out OK, never been in situation before to anticipate outcome.
13	Can you tell me what options were considered or available for managing … or does this situation fit into a standard scenario	
14	Tell me about what information you used to help you make this decision? … How did you obtain this information? … Tell me about any barriers you had obtaining this information?	Was all the information available to you to help you make the decision.
15	Can you explain to me what information you would need to make … decision?	Was all the information available to you to help you make the decision.
16	Were there any other options available at the time that should have been?	
17	Can you tell me what level of experience would be required to make this decision?	
18	For staff less experienced, describe to me what guidance/aids there are to help them make these decisions?	
	Preparedness	
19	Can you tell me a bit about the training that you/staff have to deal with this situation? … What training, knowledge, or information might have helped?	
20	Tell me what guidelines/policies there are in your unit to help manage this situation?	Any rules that applied clearly here
21	Can you tell me about any other sources of knowledge (staff, aids) you used to help you with this situation?	Application of knowledge - processing of knowledge, usable for situation, apply in correct context – was the right knowledge present for that situation.
	Communication	
22	Describe how the team communicated to each other during this time?	Communication patterns in team performance.
	Anticipation/thinking ahead	
23	Describe what you were expecting to happen?	Thinking ahead. Would also answer (did outcome fit expectation)
24	Can you explain to me what you were hoping would happen as a result of…	Imagined possible consequences of action. How were they imagining events would unfold. How did they judge they could influence the course of events?
25	What could possibly happen at this point?	What mistakes/slips were likely at this point? Don't use the word mistake when speaking to staff.

	Questions are based on/adapted from the following references: Dekker, 2014: questions 1, 2, 4, 5, 10 Hoffman et al, 1998: questions 4, 6, 11, 12, 13, 14, 19, 24 Klein, 1998: questions 3, 4, 6, 7, 9, 11, 12, 13, 19, 20, 21, 23, 24, 25 Klein et al, 2010: questions 14, 15, 16, 17, 18 Militello and Hutton, 1998: question 8 Mitchell 2013: question 12 Patel et al, 2002: questions 1, 5, 7, 8, 12, 22 Pitz and Sachs, 1984: question 10

Outcomes

The LRQ tool has been embedded into practice by the author for the past four years. Staff discussions were observed to flow better with a conversational structure. The information obtained was rich and effectively explored system or process issues, highlighting various influencing factors when standard practice was deviated from. The author reported feeling more comfortable during staff discussions as they were confident the questions they asked were system or human factors focused, rather than 'finger pointing'.

Four years into the tools use, MNSI investigators reported using the LRQ tool for planning staff discussions, real-time conversations, and when finding it hard to phrase difficult questions. MNSI investigators, new to the role, found it particularly helpful. The LRQ tool encouraged investigators to maintain an open-ended questioning approach, using an inquisitive non-blame approach that identified system issues and barriers. Investigators were able to explore nuanced information that may not have been obvious from other sources of information collected during the investigation. Using questions from the tool that explored staff's past experiences, prevented investigators from making assumptions based on staff's role or seniority. There was a desire for further support with how to ask some of the questions in the tool, along with more questions to help explore specific situations, and stress and fatigue.

Staff shared how they had been concerned about meeting with the MNSI investigator and at the end reported having had a positive experience, and greater awareness that the purpose of the meeting was to gain information to improve healthcare services rather than apportion blame. When using the LRQ tool, MNSI investigators perceived staff to show a positive response to the questions asked and were reassured and receptive, providing more information around the context of the situation.

Learning

The LQR tool is designed for 'pick and mix' use and inspiration, rather than a script, which is not recommended. Once investigators become more familiar with the LRQ tool, they will be able to integrate it into real-time staff discussions as needed, as well as use it to plan for these conversations. Reflection is always encouraged following a staff discussion to identify if there may have been a better way of phrasing a question or if the question led to staff sharing local rationality information.

Given the LRQ tool aims to explore local rationality, which in turn allows for a better understanding of how systems, processes, and human factors interact with decision-making, the tool aligns well with the Systems Engineering Initiative for Patient Safety (SEIPS) model (Carayon et al, 2006; Holden et al, 2013). However, the LRQ tool will work well alongside any investigation analysis model that allows for the 'whys' to be explored.

Limitations include the need for a formal evaluation of the tool, using recognised human factors and ergonomics methodology. The inclusion of the wider healthcare sector and other investigation industries would provide an opportunity to develop a second version of the tool in response to evaluation findings.

The use of the LRQ tool is transferable to any safety investigation industry. Human factors specialists have fed back that the tool has been positively received by investigators within the mining industry and is being shared within UK civil aviation. In healthcare, patient safety teams have fed back how the tool supports their investigations and where information obtained has led to significant changes to departmental processes and equipment. The LRQ tool can help safety investigators ask the right questions in a way that promotes psychological safety. As a result, staff share their full experience, which in turn, reduces hindsight bias by encouraging the safety investigator to focus on only the information that people had available to them at the time. Questions that promote this thinking, allow for a just culture (Eurocontrol, 2014; NHS England & NHS Improvement, 2018), while accurate findings result in effective recommendations for systemic change.

References

Carayon, P., Schoofs Hundt, A., Karsh, B. T, Gurses, A. P., Alvarado, C. J., Smith, M., and Flatley Brennan, P. (2006) Work system design for patient safety: the SEIPS model. *Quality & Safety in Health Care*, 15(Suppl 1), pp. i50–i58.

Dekker, S. (2014) *The field guide to understanding 'human error'*. 3rd edn. Surrey: Ashgate Publishing Limited.

Eurocontrol. (2014) *Systems Thinking for Safety: Ten Principles. A White Paper.* Available at: Systems Thinking for Safety: Ten Principles | SKYbrary Aviation Safety (Accessed: 24 November 2024)

Hoffman, R.R, Crandall, B. and Shadbolt, N. (1998) Use of the critical decision method to elicit expert knowledge: A case study in the methodology of cognitive task analysis. *Human Factors*, 40 (2), pp. 254-276.

Holden, R. J., Carayon, P., Gurses, A. P., Hoonakker, P., Hundt, A. S., Ozok, A. A. and Rivera-Rodriguez, A. J. (2013) SEIPS 2.0: A human factors framework for studying and improving the work of healthcare professionals and patients. *Ergonomics*, 56 (11), pp. 1-30.

Klein, G., Calderwood, R. and Clinton-Cirocco, A. (2010) Rapid decision making on the first ground: the original study plus a postscript. *Journal of cognitive engineering and decision making*, 4 (3), pp. 186-209.

Klein, G.A (1998) *Sources of power: How people make decisions*. Cambridge: MIT Press.

Militello, L.G. and Hutton, R.J.B. (1998) Applied cognitive task analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands. *Ergonomics*, 41 (11), pp. 1618-1641.

Mitchell, P. (ed.) (2013) *Human factors for healthcare: Trainers manual*. Cove: Swan and Horn.

NHS England & NHS Improvement (2018) *A Just Culture Guide*. Available at: https://www.england.nhs.uk/wp-content/uploads/2021/02/NHS_0932_JC_Poster_A3.pdf (Accessed 12 February 2025).

Patel, V.L., Kaufman, D.R. and Arocha, A.F. (2002) Emerging paradigms of cognition in medical decision-making. *Journal of biomedical informatics*, 35, pp. 52-75.

Pitz, G.F. and Sachs, N.J. (1984) Judgement and decision: theory and application. *Annual review of psychology*, 35, pp. 139-163.

Wu, A.W. and Steckelberg, R.C. (2012) Medical error, incident investigation and the second victim: doing better but feeling worse? *BMJ Quality and Safety*, 21(4), pp. 267-270.

Specific Heuristics for Smartwatch Usability Evaluation: Development, Validation and Comparison

Yiyao Li, Maria Richart & Setia Hermawati

Human Factors Research Group, University of Nottingham, UK

SUMMARY

This study examined the effectiveness of specialised heuristics for smartwatches, focusing on unique usability challenges of wearable interfaces. The study used a combination of literature review, focus groups, and expert evaluation to develop ten heuristics for smartwatch interfaces. Subsequently, an empirical study of five typical tasks on smartwatch was conducted, with three groups of participants using Nielsen's heuristics, specialised heuristics, and user testing to collect data. By combining theoretical development and empirical validation, this research proposed a framework for adapting and validating heuristic evaluations to emerging wearable technologies. Our findings showed that the heuristics set for smartwatch was more effective than general heuristics due to its better accuracy in anticipating notable usability concerns.

KEYWORDS

Usability evaluation, heuristic evaluation, wearable technology, human-computer interaction (HCI)

Introduction

Since the invention of the first smartwatch in the early 90s, the interface and functionality of the smartwatch have continued to improve and adapt to the needs of users (Radnejad, Ziolkowski, & Osiyevskyy, 2020). As wearable technologies, particularly smartwatches, become increasingly integrated into daily life, their unique design challenges demand a shift in usability evaluation methods (Stefana et al., 2021). The most popular and standardised heuristics used were developed by Nielsen and Molich in 1990 and then later finalised by Nielsen in 1994, incorporating feedback from expert evaluators (Nielsen, 1994a). However, traditional heuristic evaluation frameworks, such as Nielsen's heuristics, were developed for desktop interfaces and lack consideration for the constrained screen size, diverse input methods, and ergonomic considerations of wearable devices (Park, Jeong, & Kim, 2020). To address this gap, we developed and validated a set of smartwatch-specific usability heuristics. This study aims to present the development process and validation of these heuristics by comparing them with Nielsen's framework to assess their effectiveness in identifying real-world usability issues in smartwatch interfaces.

This research is particularly relevant to the field of human factors and ergonomics, where understanding the specific usability needs of small-screen wearable devices is crucial (Darmwal, 2015). By assessing the effectiveness, rigor, and relevance of both general and device-specific heuristics, this study contributes to the ongoing development of practical, specialised tools for the evaluation of wearable technology interfaces. This approach seeks to empower user experience practitioners with targeted heuristics that can be widely adopted in the design and evaluation of next-generation wearable devices.

Methodology

This study integrated two research phases which followed the method proposed by Hermawati and Lawson (2015). In phase one, we conducted a comprehensive literature review, followed by three user focus groups with three participants each, to understand smartwatch-specific usability challenges. The focus group sessions were conducted online. The moderator created an informal communication experience for users during these sessions while guiding them to discuss their attitudes, beliefs, and perceptions about the usability of smartwatches. Each session was directed by a set of pre-prepared questions that focused on the usability of smartwatches, particular issues when using smartwatches, and recommendations for enhancing the usability of smartwatches. The utilization of focus groups to gather genuine user experience and feedback served to enhance the applicability of heuristics, moving beyond a reliance solely on literature-based approaches. The outcomes of the focus group discussions were recorded, systematically coded, and then comprehensively compared and analysed in conjunction with the findings from the literature review. This informed the creation of an initial set of heuristics. The last step in phase one was to present these initial set of heuristics to three Human-Computer Interaction experts so they could review and provide feedback on their relevance and clarity. Three experts were invited by email, selected based on their expertise in human-computer interaction and human factors. The experts assessed each usability heuristic on its applicability, consistency, understandability, completeness, redundancy, scalability, and terminology usage. They were also invited to suggest modifications or draw attention to any omissions. The usability heuristics for smartwatches were then refined and finalised based on the feedback from experts. This step served to enhance the clarity and generality of the proposed heuristics.

In phase two, we conducted a validation study with three participant groups, each consisting of five participants. The first group used Nielsen's heuristics, the second group applied the smartwatch-specific heuristics, and the third group conducted user testing to establish a baseline of real-world usability issues. For testing purposes, a single type of smartwatch with a single version of that interface needed to be used to standardise the experiment. All groups evaluated the usability of the Apple Watch Series 9 across five common tasks: check daily activity, record an outdoor walk, check the weather, play music from watch using Bluetooth, and check past notifications. Before the heuristic evaluation could commence, an instructional guide containing explanations of all five tasks that included pictures of the smartwatch interface and descriptions needed to be made. This was what the participants in both the Nielsen's heuristics evaluation and the smartwatch heuristic evaluation would use to examine and determine usability issues in the interface. User testing study was done with one of the researchers who observed and recorded each participant's interaction with the smartwatch during the tasks. Each of the five participants recruited for this group had an Apple Series 9 Smartwatch on which they performed each requested task. The usability issues were then compiled and coded to aid comparison between the groups. The results were then analysed based on three key metrics: *thoroughness*, which pertains to the capacity to identify pertinent real-world issues; *validity*, reflecting the accuracy in anticipating notable usability concerns; and *effectiveness*, signifying the overall achievement in conducting the usability evaluation (Hartson, Andre, & Williges, 2003). In addition, the statistical analysis was further conducted in SPSS across the various groups, encompassing a one-way ANOVA to compare the number of violations and a Kruskal-Wallis test to compare the median severity ratings. Lastly, each set of ten and the corresponding number of real-world problems each heuristic correctly identified were analysed with one-way ANOVA in SPSS. The goal was to see which heuristics within each were best and worst at identifying real-world problems which was the basis of which heuristics to focus on when making recommendations to improve the entire smartwatch heuristic set. The steps of Heuristic Comparison Study are shown in Figure 1.

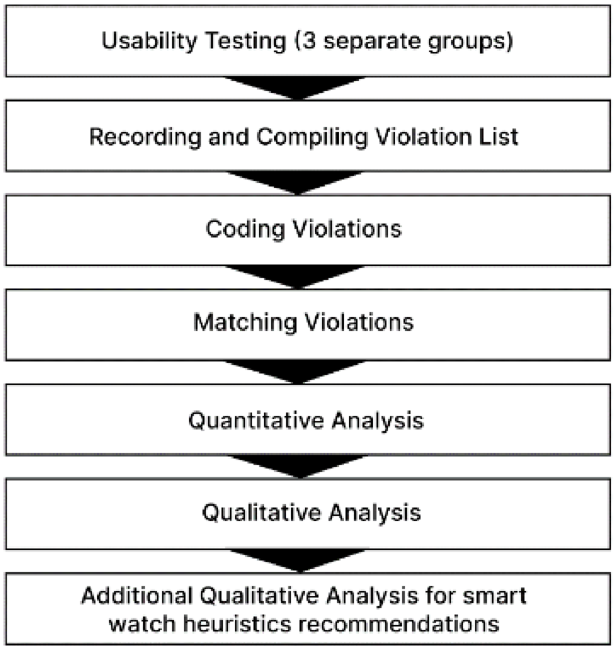


Figure 1: Heuristic Comparison Study

Main findings

Ten heuristics were identified from phase one, which included: “Visibility of System Status”, “Readability”, “Simplicity and Intuitiveness”, “Contextual Feedback”, “Minimised Operational Effort”, “Fault Tolerance”, “Consistency”, “Input Method Control”, “Interoperability” and “Ergonomic Design”. Each heuristic was devised to address aspects of smartwatch usability, including system state feedback, readability, effective interaction, and comfort. These heuristics were selected based on their alignment with user needs and pain points, as evidenced by focus group research and confirmed by existing literature. For instance, due to the small screen size of smartwatches, the focus groups repeatedly highlighted the significance of visibility and readability. Similarly, interoperability was heavily favoured in the literature, stressing the increasing requirement for seamless integration of devices. It should be noted that the usage experience issues that were repeatedly mentioned in the group discussions can be considered common user pain points and thus should be prioritised as focus areas for heuristics, such as improving system status visibility, gesture interaction, and screen readability. The literature review and the results of the group discussions were used to determine the prioritisation of heuristics. In the ranking of the heuristics, the issues that mattered most to users were prioritised, for example, “Visibility of system status” was placed first.

Results of phase two showed that 66, 54 and 42 usability issues were identified from Nielsen’s and smartwatch heuristics set, and user testing, respectively. When the codes of the unique violations of the Nielsen’s heuristics evaluation and the smartwatch heuristic evaluation were matched to the user testing results, it was found that the Nielsen’s heuristics predicted 27 real-world violations, and the smartwatch heuristics predicted 25. To calculate the thoroughness value for each, the number of correctly predicted user violations in each heuristic evaluation was divided by the 42 violations discovered in user testing. The calculation showed that Nielsen’s heuristics had a thoroughness value of 0.643, while the smartwatch heuristics’ value was 0.595. In other words, Nielsen’s heuristics identified 59.5% of the total real-world problems, while the smartwatch heuristics identified 57.1%. To calculate the validity value, the number of correctly predicted violations for each heuristic set was divided by the total number of violations identified by that set. The calculation showed that Nielsen’s heuristics had a validity value of 0.409, while the smartwatch heuristics' validity value was 0.463. This meant that 40.9% of the 66 usability issues predicted by Nielsen's set were real-world issues, while 46.3% of the 54 usability issues identified by the smartwatch heuristics were real-world issues. Finally, the thoroughness and validity value were multiplied to determine the overall effectiveness value of each heuristic set. The results showed that the effectiveness value of Nielsen's heuristics and smartwatch heuristics were 0.263 and 0.275 respectively. The overall calculation results are presented in Table 1.

Table 1: Calculation results of three indicators for each heuristic set

Heuristic evaluation methods	**Thoroughness value**	**Validity value**	**Effectiveness value**
Nielsen's heuristics	0.643	0.409	0.263
Smartwatch heuristics	0.595	0.463	0.275

Additionally, the one-way ANOVA results showed that there was a significant difference in the number of violations each usability evaluation method found ($p < 0.001$). Post hoc tests revealed that there was a significant difference between user testing and Nielsen's heuristics evaluation ($p=0.003$) and the user testing and the smartwatch heuristics evaluation ($p=0.004$). However, there were no significant differences between two heuristic evaluations ($p=0.546$). The Kruskal-Wallis test found a significant difference in the severity ratings of the three methods ($p=0.047$). Post hoc tests showed that there was a significant difference between the user testing and Nielsen's heuristics evaluation ($p=0.016$), but not between the user testing and the smartwatch heuristics evaluation ($p=0.057$) or between the two heuristic evaluation methods ($p=0.974$). Furthermore, the correctly predicted usability violations for each heuristic were analysed using one-way ANOVA. The statistical analysis showed that the "flexibility and efficiency of use" heuristic from Nielsen's set was the most effective in discovering navigation-related violations, prompting its inclusion in the smartwatch heuristics set. In contrast, "Consistency and Standards," "Error Prevention," and "Help Users Recognize, Diagnose, and Recover from Errors" were the least effective. Among the smartwatch heuristics, "Fault Tolerance" and "Ergonomic Design" were identified as the least effective and require improvement, while "Visibility of System Status" and "Readability" emerged as the most effective heuristics.

Discussion

Analysing the thoroughness, validity, and effectiveness of each set helped determine which was better suited for heuristic evaluation of smartwatch interfaces. Based on the results of the thoroughness value, both the Nielsen's and smartwatch set had decently high values as they were able to find more than half of the real-world issues—0.643 (64.3%) and 0.595 (59.5%) respectively. This suggests that they were often able to find real-world usability issues at a very similar rate. However, it is not just thoroughness that determines which set is more effective, and therefore better at evaluating smartwatch design, but also their validity value. This value represents the accuracy of each set of tests and their ability to find actual real-world violations instead of many insignificant violations by finding the percentage of correctly predicted violations out of all violations found. The smartwatch heuristics set exhibited a higher validity value of 0.463 compared to Nielsen's 0.409, indicating a superior ability to pinpoint relevant usability violations. Finally, with the above values the overall effectiveness of each set was calculated. It was found that the effectiveness of Nielsen's heuristics had a value of 0.263, while the smartwatch set had a value of 0.275. With these values it can be concluded that the smartwatch heuristic set was more effective at predicting real world usability issues in a heuristic evaluation of smartwatches than Nielsen's set.

Therefore, in this case, tailoring a set of heuristics to the specific technology of a smartwatch proved to be better than using a more generic set. General heuristics, while useful, may overlook critical design issues when applied to unique technologies like wearables. Therefore, custom heuristics should be used to improve usability evaluation outcomes by accounting for device unique features and interaction styles. For instance, the concepts of "Readability" and "Simplicity and Intuitiveness" are not mere derivatives of broad usability principles, but rather, they tackle distinct issues that arise from the constrained screen size and diverse input mechanisms seen on smartwatches. Within the distinctive framework of the gadget, this heuristic has the potential to provide significant assistance to designers. Similarly, the heuristic known as "Input method control" emphasizes the need of smartwatches being able to accommodate many types of input, such as touch, speech, physical buttons, and gestures. This is necessary because smartwatches are frequently utilized in diverse situations, ranging from fitness activities like running on a track to attending meetings and checking alerts. In contrast to smartphones and computers, which predominantly rely on touch and keyboard as primary input modalities, smartwatches necessitate a higher degree of adaptability owing to their multifaceted applications. Moreover, the heuristic of "Interoperability" effectively addresses the persistent challenge of multidevice interaction within the smartwatch industry since smartwatches frequently function within a network of interconnected smart devices.

To further analyse whether the smartwatch heuristic set was better suited for evaluate smartwatch design than the Nielsen's set, the number of violations each found as well as the median severity participants gave to each were analysed. A significant difference found in the number of violations of each heuristic set and user testing (Nielsen's: $p = 0.003$, Smartwatch: $p = 0.004$) but not between the heuristic sets ($p=0.546$). This supports previous research including that in the paper "Number of People Required for Usability Testing" that heuristic evaluations find a significant amount more of violations than user testing (Hwang & Salvendy, 2010). The results of the Kruskal-Wallis test relating to the median of the violation severity rating showed that there was a significant difference between the severity rating of violations in user testing and Nielsen's ($p = 0.016$). In contrast, there was no significant difference in the median severity rating of violations between smartwatch heuristics set and user testing ($p = 0.57$). This suggests that the smartwatch heuristic set was better at estimating a more accurate severity than the Nielsen's set. This is a favourable attribute of smartwatch heuristics set, which also proves it to be better than the Nielsen's set.

The results from the comparison of individual heuristic in each set showed which heuristic was better at finding the problems that existed within the smartwatch design and can be used to improve the smartwatch heuristic set. Regarding Nielsen's heuristics, a statistically significant difference was observed between the heuristic "Flexibility and Efficiency of Use" and three less effective heuristics: "Consistency and Standards," "Error Prevention," and "Help Users Recognize, Diagnose, and Recover from Errors." The superior effectiveness of "Flexibility and Efficiency of Use" suggests its relevance in addressing smartwatch usability challenges, particularly in navigating small-screen interfaces with shortcuts and streamlined actions. Given the absence of an equivalent heuristic in the smartwatch-specific set, it is recommended for inclusion. The smartwatch heuristic evaluation revealed considerable variance in the accuracy of usability issue identification across heuristics, with some failing to detect any issues. "Visibility of System Status" and "Readability" emerged as the most effective heuristics, significantly outperforming others such as "Contextual Feedback," "Input Method Control," and "Fault Tolerance." The effectiveness of "Visibility of System Status" and "Readability" can be attributed to their highly visual nature, allowing usability issues to be identified even through static images of the interface. This helps provide understanding for why the heuristics of "Input M ethod Control", "Contextual Feedback", "Ergonomic Design", and "Interoperability" were significantly less effective as the real-life usability issues went beyond what could be found from looking at an image. "Ergonomic Design" is very relevant as a watch's small screen is worn and operated on a user's wrist. If this heuristic is to work within heuristic evaluation, it should be redefined to frame this as how intuitively placed buttons and dials are and how comfortable it is to physically access these functions. "Fault Tolerance" was found to be ineffective in heuristic evaluation despite user testing identifying numerous related issues. The limitations of this heuristic can be addressed by splitting it into two distinct components: "Fault Prevention" and "Fault Recovery." "Fault Prevention" would guide evaluators to identify conditions likely to cause usability errors, such as a lack of confirmation for consequential or ambiguous actions. "Fault Recovery" would focus on mechanisms enabling users to recognize and correct errors efficiently. By explicitly defining these conditions, evaluators can more effectively detect usability violations. Table 2 lists the refined smartwatch heuristics based on these recommendations.

Table 2: Refined usability heuristics for smartwatches

No.	Usability heuristic	Description
1	Visibility of system status	Smartwatches should provide immediate feedback about status including Bluetooth and battery so that users have certainty to engage in tasks on the device.
2	Readability	Information on the interface must be appropriate in size, colour, position, and contrast to provide suitable communication to users in a variety of settings.
3	Simplicity and intuitiveness	Information conveyed should be straightforward and intuitive so that crucial details can be relayed in a rapid and succinct manor.
4	Contextual feedback	Confirmation about actions taken including haptic, visual, and audio feedback should be timely and purposeful to guide the user.
5	Minimised operational effort	Processes of interactions and task completion should require only necessary steps and should be able to be done with minimal exertion.
6	Fault prevention	Processes of interactions and task completion should require only necessary steps and should be able to be done with minimal exertion.
7	Fault recovery	The design should have a strong protocol for recovery especially in critical times and provide guidance to easily get back to the correct task process including error messages and suggested solutions.
8	Consistency	A cohesive set of design elements and language need to be present across all components of the interface so that a user can apply their understanding of this across the overall design interface.
9	Input method control	Users should be able to interact with the interface in multiple ways including touch, buttons, voice commands, and gestures to be able enhance the device's accessibility and usability.
10	Interoperability	Seamless and quick integration between the smartwatch and other devices such as mobile phones and computers should be present so that information transmitted between them enhances the user's experience.
11	Ergonomic design	The design of the smartwatch, including how intuitively the buttons and other interactive features are placed, should

		prioritise physical comfort for the purposes of the activities of the user including prolonged durations of usage.
12	Flexibility and efficiency of use	Short cuts and customisable features are available to users to decrease the time it takes to perform tasks.

Conclusion and future work

The development of usability heuristics that are specifically tailored for smartwatches is a crucial undertaking to enhance the overall usability of these technological devices. The 12 usability heuristics that emerged from this study encompass a diverse range of crucial factors for assessing the usability of smartwatches. Significantly, these heuristics not only encompass overarching concepts of usability, but also derive from the distinct challenges and possibilities posed by smartwatches. Every principle was derived from the analysis of user requirements and challenges, aiming to address distinct issues pertaining to the usability of smartwatches. This study, which compared Nielsen's heuristics with a smartwatch-specific heuristic set, revealed that while Nielsen's set identified the highest number of issues overall, its effectiveness was comparatively lower. In contrast, the smartwatch heuristic evaluation detected slightly fewer issues but demonstrated a higher accuracy rate and greater overall effectiveness. In other words, the set of smartwatch heuristics designed specifically for smartwatches and their features proved to be better and more effective when used in the heuristic evaluation. This finding, along with the recommendations proposed in this study, suggests that adapting heuristics to specific technologies can be more effective than applying a standard heuristic set such as Nielsen's. These insights may also have broader implications for heuristic evaluation in other domains.

While the smartwatch heuristics proved to be more effective in this study, there is still room for improvement. To further refine the heuristic set for smartwatch design, the next step should be to validate the newly refined heuristics based on this study's recommendations. Another way to expand this research is to include a wider range of smartwatch brands and models, as this study was limited to testing the Apple Watch Series 9. In doing so, a larger variety of smartwatch designs could be analysed and increase the understanding of what heuristics and descriptions are needed to capture as many usability issues as possible. In conclusion, this study confirms that although general heuristic evaluation can be applied to smartwatches, heuristic evaluation optimized for smartwatches is more effective. This study not only provides preliminary verification, but also proposes specific optimization directions, providing a more refined methodology for future usability evaluation. Future research can further expand the sample size and optimize certain inefficient heuristic principles to further improve the evaluation effect.

References

Darmwal, R. (2015). Wrist Wars: Smart Watches vs Traditional Watches. Telecom Business Review, 8(1), p. 69.

Hermawati, S. & Lawson, G. (2015). A user-centric methodology to establish usability heuristics for specific domains. In Proceedings of the International Conference on Ergonomics & Human Factors, pp. 80-85.

Hwang, W., & Salvendy, G. (2010). Number of people required for usability evaluation. Communications of the ACM, 53(5), 130–133.

Hartson, H.R., Andre, T.S. & Williges, R.C. (2003). Criteria For Evaluating Usability Evaluation Methods. International Journal of Human–Computer Interaction, 15(1), pp. 145–181.

Nielsen, J. (1994a). Enhancing the explanatory power of usability heuristics. Proc. ACM CHI'94 Conf. (Boston, MA, April 24-28), 152-158.

Park, K., Jeong, M. & Kim, K. (2020). Usability evaluation of menu interfaces for smartwatches. The Journal of computer information systems, 60(2), pp. 156–165.
Radnejad, A. B., Ziolkowski, M. F., & Osiyevskyy, O. (2020). Design thinking and radical innovation: Enter the smartwatch. Journal of Business Strategy, 42(5), 332–342.
Stefana, E., Marciano, F., Rossi, D., Cocca, P., & Tomasoni, G. (2021). Wearable Devices for Ergonomics: A Systematic Literature Review. Sensors, 21(3), 777.

Exploring Human Performance in Mako-Assisted Hip Replacement Surgeries

Jasper Vermeulen[1,2], Prof Glenda Caldwell[1,2], A/Prof Müge Belek Fialho Teixeira[1,2], Alan Burden[1,2], Matthias Guertler[1,3]

[1]Australian Centre for Cobotics, AU [2]Queensland University of Technology, AU [3]University of Technology Sydney, AU

SUMMARY

This study investigates the human factors shaping surgical workflows in Mako-assisted Total Hip Arthroplasty surgeries, particularly emphasising non-technical skills such as communication, teamwork, and situational awareness. Using video analysis, we examine interactions between surgical team members and the Mako system, identifying challenges and opportunities that enhance future collaborations.

KEYWORDS

Mako, Human Factors, Non-Technical Skills, Total Hip Arthroplasty, Robot-assisted surgery, Human-Robot Collaboration

Introduction

Human-Robot Collaboration (HRC) is an exciting prospect as it encapsulates using robotic technologies to enhance human performance (Burden et al., 2022; Kopp et al., 2019). HRC often builds on the idea of human-robot symbiosis, using the respective strengths of humans and robots to achieve greater outcomes (Wang et al., 2019). To achieve HRC, robots are often designed to perform closely together with humans on the same task. Specifically, robots are designed to alleviate physical ergonomics; reduce physical strain and improve posture. However, the recent work of Vermeulen et al. (2024) draws attention to a potential weighing of human factors in HRC. Their literature review showcases how HRC can improve physical ergonomics but might also induce negative cognitive consequences (Vermeulen et al., 2024). An example of this is workers experiencing stress from having to work closely with robots (Faccio et al., 2023). Furthermore, their literature review highlights that human factors studies in HRC play a crucial role in this field as most of the literature historically has been techno-centric, neglecting the role of the humans collaborate with the robot (Vermeulen et al., 2024). The handful of human factors studies in HRC tend to study humans in laboratory or experimental settings. Therefore, a need has been identified to understand human factors in HRC tasks conducted in real-world contexts.

Theoretical Background

Human factors studies on HRC have predominantly focused on one specific robotic system: the da Vinci Surgical System (Catchpole et al., 2019, 2024). The da Vinci is a teleoperated robot, which means the surgeon positions themselves at a control console to perform surgical tasks. Cheatle et al. (2019) and Pelikan et al. (2018) investigated the human factors in da Vinci-assisted surgeries using observations and interviews with the different surgical team members. Their findings highlight some of the challenges surgical teams face due to the insertion of robotic assistance. In this case, with the da Vinci, communication and teamwork challenges arise due to the surgeon being physically distanced from the rest of the surgical team and the patient. Such a human factors exploration has yet to be conducted for Mako-assisted surgeries.

Methodology

We utilised Atlas.ti, a qualitative data analysis software, to systematically analyse the video content. Our approach comprised the following steps:
(a) we identified the distinct tasks involved in THA surgeries, with particular attention to those requiring interaction with the Mako;
(b) we mapped these tasks across the selected videos to establish general workflows for Mako-assisted procedures based on in-situ practices;
(c) we coded the interactions and communication between the surgeon and other surgical team members;
(d) we grouped the resulting codes into themes to uncover the key human factors affecting the workflow in Mako-assisted THA surgeries.

Through this analysis, we aim to understand better how human factors influence surgical processes and identify opportunities to enhance collaborative efficiency in robotic-assisted surgery.

Results

Our findings reveal distinct communication patterns that emerge during Mako-assisted THA surgeries. The Mako is pivotal in facilitating effective communication among surgical team members. This communication encompasses visual, auditory, and haptic modalities, with the latter exclusive to the surgeon. The Mako Product Specialist (MPS) and the scrub nurse significantly support the surgeon's successful collaboration with Mako. The MPS's technical expertise ensures that the surgeon can effectively operate the Mako system, often guiding the surgeon through the software interface while following the surgeon's directives. Scrub nurses complement this support by assisting the MPS with tasks such as preparing and positioning the robot and ensuring the appropriate instruments are promptly provided to the surgeon. Our analysis highlights the critical role of effective communication and teamwork in the success of Mako-assisted THA surgeries. These human factors enhance operational efficiency and reduce surgical duration, ultimately benefiting both the surgical team and patient outcomes.

A reduced surgical duration offers significant benefits for the surgeon, particularly in THA surgeries, where certain tasks, such as bone reaming, require operating in ergonomically challenging positions. Due to the specific angles needed, surgeons often lift their arms above shoulder level, which can increase the risk of strain or injury if sustained over extended periods. One surgeon in our video observations used a stool to elevate themselves to alleviate this issue, achieving a slightly improved—though still suboptimal—ergonomic position.

In addition to ergonomics, the spatial layout of the operating room and the positioning of surgical team members play a crucial role in the workflow of THA surgeries. For the Mako robotic system to function effectively, its sensors must maintain an unobstructed view of the robotic arm and the patient. However, the often crowded and constrained operating theatre can lead to instances where team members inadvertently block the robot's field of vision, causing workflow disruptions. Optimising the movement and positioning of surgical actors, such as the MPS and scrub nurse, could address these challenges. By analysing these movements, opportunities for improved spatial organization and positioning strategies can be identified, ultimately enhancing HRC and promoting a more efficient workflow in Mako-assisted THA surgeries.

Conclusion

This paper holds significance for the human factors community as it represents a pioneering exploration of a novel type of robot that has not been previously studied within this domain. However, the study does have limitations. The reliance on video recordings may only partially capture the complexity and nuances of these surgical procedures. Moreover, the videos included best use cases of the system, potentially omitting crucial workflow disruptions. Future research would benefit from incorporating in-person observations to provide a more comprehensive understanding and richer insights into the dynamics of Mako-assisted surgeries.

References

Burden, A. G., Caldwell, G. A., & Guertler, M. R. (2022). Towards human–robot collaboration in construction: current cobot trends and forecasts. *Construction Robotics*, *6*(3), 209–220.

Catchpole, K., Bisantz, A., Hallbeck, M. S., Weigl, M., Randell, R., Kossack, M., & Anger, J. T. (2019). Human factors in robotic assisted surgery: Lessons from studies "in the Wild." *Applied Ergonomics*, *78*, 270–276.

Catchpole, K., Cohen, T., Alfred, M., Lawton, S., Kanji, F., Shouhed, D., Nemeth, L., & Anger, J. (2024). Human factors integration in robotic surgery. *Human Factors*, *66*(3), 683–700.

Cheatle, A., Pelikan, H., Jung, M., & Jackson, S. (2019). Sensing (Co)operations: Articulation and Compensation in the Robotic Operating Room. *Proc. ACM Hum.-Comput. Interact.*, *3*(CSCW), 1–26.

Dretakis, K., Raptis, K., & Koutserimpas, C. (2024). The use of the robotic arm-assisted system (MAKO) for hip revision surgery. *The Archives of Bone and Joint Surgery*, *12*(8), 608–611.

Faccio, M., Granata, I., Menini, A., Milanese, M., Rossato, C., Bottin, M., Minto, R., Pluchino, P., Gamberini, L., Boschetti, G., & Rosati, G. (2023). Human factors in cobot era: a review of modern production systems features. *Journal of Intelligent Manufacturing*, *34*(1), 85–106.

Guo, X., Wang, D., Li, J., & Zhang, H. (2023). Global research status and trends in orthopaedic surgical robotics: a bibliometric and visualisation analysis study. *Journal of Robotic Surgery*, *17*(4), 1743–1756.

Jasper Vermeulen, Glenda Caldwell, Müge Belek Fialho Teixeira, Alan Burden, Matthias Guertler. (2024). To Safety and Beyond! A Scoping Review of Human Factors Enriching the Design of Human-Robot Collaboration. *Interaction Design and Architecture(s) Journal*, *61*(Designing for People in Human-Robot Collaboration), 42–65.

Kopp, R., Dhondt, S., Hirsch-Kreinsen, H., Kohlgrüber, M., & Preenen, P. (2019). Sociotechnical perspectives on digitalisation and Industry 4.0. *International Journal of Technology Transfer and Commercialisation*, *16*(3), 290–309.

Pelikan, H. R. M., Cheatle, A., Jung, M. F., & Jackson, S. J. (2018). Operating at a distance - how a teleoperated surgical robot reconfigures teamwork in the operating room. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 1–28.

Subramanian, P., Wainwright, T. W., Bahadori, S., & Middleton, R. G. (2019). A review of the evolution of robotic-assisted total hip arthroplasty. *Hip International: The Journal of Clinical and Experimental Research on Hip Pathology and Therapy*, *29*(3), 232–238.

Wang, L., Gao, R., Váncza, J., Krüger, J., Wang, X. V., Makris, S., & Chryssolouris, G. (2019). Symbiotic human-robot collaborative assembly. *CIRP Annals*, *68*(2), 701–726.

Wu, X.-D., Zhou, Y., Shao, H., Yang, D., Guo, S.-J., & Huang, W. (2023). Robotic-assisted revision total joint arthroplasty: a state-of-the-art scoping review. *EFORT Open Reviews*, *8*(1), 18–25.

Understanding the Human Factors Related to Unrecognised Oesophageal Intubation Using the SEIPS Framework

Melody Chen[1,2], B. L. William Wong[2]

[1]Te Whatu Ora, Health New Zealand, [2]Auckland University of Technology, New Zealand

SUMMARY

Unrecognised oesophageal intubation (UOI) is a medical procedural error in which a breathing tube is mistakenly placed into the oesophagus and not promptly identified or addressed. A literature review on the human factors relating to UOI was performed and findings presented in the Systems Engineering Initiative for Patient Safety (SEIPS) framework. Key themes relating to issues around intubation equipment factors were used to inform the design of a cognitive aid to improve human performance in this clinical context. A semantically meaningful tray consisting of images and uniquely sized slots was created and sought to address technical and non-technical human factors identified in adverse event reports and coronial cases of UOI.

KEYWORDS

SEIPS model, Human factors, Healthcare, Airway Management, Unrecognised Oesophageal Intubation

Introduction

A patient who is critically ill in an intensive care unit (ICU) may require ventilation through an artificial airway, which involves the insertion of an endotracheal tube (ETT) into the trachea. It is possible that the ETT can be unintentionally inserted into the oesophagus, and further, go unrecognised by the intubator. Unrecognised oesophageal intubation (UOI) puts the patient at risk of immediate death or major harm (Baker et al., 2022). Episodes of UOI has been reported in healthcare organisations around the world, however the term 'unrecognised oesophageal intubation' was only recently defined in 2022 as "unintended placement or migration of a tracheal tube into the oesophagus, that is not promptly identified and addressed" (Chrimes et al., 2022).

The incidence of recognised OI was reported in a recent international study as 1 in 18 cases of emergency intubations of critically ill patients in the ICU (Russotto et al., 2021). It has been reported to occur in both routine and challenging airways (Holland et al., 1993), with both experienced and inexperienced clinicians and in all parts of the hospital including the ICU, emergency department (ED) and operating rooms (OR) (Chrimes et al., 2022). Currently, there is no structured method for estimating the true incidence rate of UOI and occurrences are mostly highlighted by coronial or media reports, meaning that it is likely more cases exist than are recorded publicly (Mann et al., 2023).

A literature review was performed to understand how UOI occurs at a macro-level in sociotechnical systems such as the hospital. The findings were framed in the Systems Engineering Initiative for Patient Safety (SEIPS) model to clearly outline work system factors that increase the risk of poor patient safety outcomes. Key findings derived from the SEIPS analysis involving access to important equipment such as capnography, were highlighted. An analysis of two New Zealand cases from 2024 (HDC, 2024), provided an understanding of how UOI occurs at a micro-level. Issues around capnography were again highlighted in these cases and linked to findings from the SEIPS analysis. These findings address only a portion of the wider body of work related to prevention of UOI but have directed the design of a cognitive aid to address technical and non-technical factors identified in adverse event reports and coronial cases. A project was initiated to design a semantically meaningful tray consisting of images and uniquely sized slots to improve operator performance during equipment preparation. Clinicians who were directly involved with the recent UOI case were involved in the design of the tray, discussed herein.

An evaluation of the new tray with nurses to understand the usability of this tray within the context of the cardiovascular intensive care unit (CVICU) in simulated intubation scenarios. The completeness of equipment preparation, time taken to prepare equipment and ability to identify missing equipment were measured to inform the effectiveness of this aid.

Background

In 2011, the 4th National Audit Project (NAP4), 'Major complications of airway management in the United Kingdom' identified nine cases of UOI, leading to six deaths which accounted for 18% of the total 33 deaths relating to airway management (Cook et al., 2011). This prompted a national campaign: 'No trace = wrong place' to highlight the importance of correct use of capnography (method to measure continuous exhaled CO_2) and human factors relating to this situation (Baker et al., 2022). Unrecognised oesophageal intubation was initially included on the Never Events list in 2018 as it is considered a preventable event through adherence to published guidelines and strong systemic protective barriers at a national level (NHS Improvement, 2018). It has since been suspended from the Never Events list in 2018 pending further clarification. The Royal College of Anaesthetists continue to emphasise that the adherence to proper monitoring of exhaled CO_2 (using waveform capnography) and its correct interpretation as the gold standard practice for confirming ETT placement (Chrimes et al., 2022). In 2022, the Project for Universal Management of Airways (PUMA) guidelines were published and is the first guideline dedicated to preventing UOI. It provides guidance on technical aspects of task performance, equipment choice and decision-making, and addresses the non-technical aspects such as human factors (Chrimes et al., 2022). All major airway societies have endorsed this guideline (Mann et al., 2023).

Literature review

Methodology

A literature review was performed to capture the interacting elements of the sociotechnical system surrounding this aspect of airway management. Articles were sourced from a combination of manual database searching and Elicit AI, an artificial intelligence research tool that uses large language models to search for relevant articles within the Semantic Scholar database (Elicit, 2024). Elicit AI was used to source 30 articles relevant to the research question "What are the human factors related to unrecognised oesophageal intubation?". Elicit AI was used to generate insights from these articles, and this was coupled with manual reading of the same papers to validate insights. To counteract the effect of learning bias, half the papers were analysed by Elicit AI first and the remaining half analysed by manual analysis first. Each insight generated by Elicit AI was checked manually for correctness and those that were deemed inaccurate due to AI related errors were removed (hallucination and referencing errors). Bibliographies from relevant articles were searched for additional research of interest. Searches of MEDLINE, PUBMED and SCOPUS were performed with the initial keywords: Unrecognise*, delay*, oesophag*, esophag*, intubation to capture further relevant articles. A total of 45 articles were included in the review.

Thematic analysis was used to group findings into higher level themes and were then organised into the Systems Engineering Initiative for Patient Safety (SEIPS) framework. The SEIPS model, first published in 2006 by Caryon et al. (2006) and has been widely used in healthcare to frame the design and research of improvement initiatives relating to patient safety. The model describes the interactions between the work system, processes and outcomes. The work system is comprised of five elements: person, task, tools and technology, environment and organisation (Figure 1).

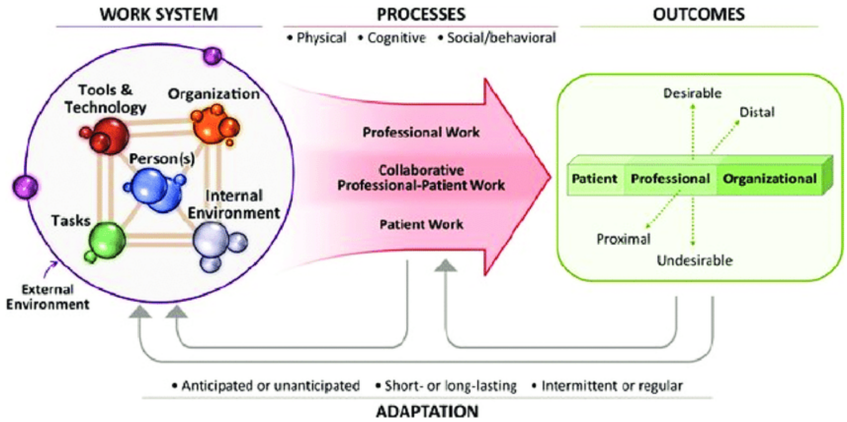


Figure 1: SEIPS model diagram

SEIPS 2.0 (Holden et al., 2013) describes the concept of configuration, which captures the dynamic nature of a system, making it possible to depict the performance of a system based on how the work system is configured. This implies that the same system, configured differently, can result in both desirable and undesirable outcomes. In this context, an *undesirabl*e outcome is unrecognised oesophageal intubation. Using evidence in literature relating to the cases of UOI, factors relating to how the person, technology, task, environment and organisation contributed to undesirable outcomes were analysed and a snapshot of the work system configuration that likely leads to poor system performance was captured.

Work system factors

Factors of the *person or team* in the context of the SEIPS model describe aspects such as knowledge deficits, anchoring or fixation and cognitive biases. **Knowledge deficits** included misinterpretation of the capnography trace or failure to acknowledge the absence of end tidal carbon dioxide ($ETCO_2$) as a sign for incorrect tube placement, has been reported as one of the leading reasons for UOI (Endlich et al., 2024). **Anchoring or fixation** to an initial diagnosis can lead to tunnel vision and in many cases of UOI, coroners reported that practitioners commonly misattributed the absence of sustained exhaled CO2 to bronchospasm (Sakles et al., 2023) continuing to treat the wrong problem and delaying the identification of OI. Clinical assessments such as listening for breath sounds are prone to subjectivity (Jafferji et al., 2019) as air flowing through the tube into the stomach can be misinterpreted as breath sounds in the lungs (Honardar et al., 2017). Subjective tests are likely to reaffirm their initial diagnosis via confirmation bias.

Technology factors involve the lack of availability of recommended gold standard equipment and delayed access to equipment. The PUMA guidelines highlight the current **'gold standard' equipment** to be used in all airway intubations. Video laryngoscopy is recommended over direct laryngoscopy as it allows the larynx view to be projected onto electronic monitors visible to more team members, not just the intubator (Kelly & Cook, 2022). It is thought to reduce the risk of fixation bias by sharing the view with multiple people in the room. The gold standard for assessing ETT placement is with continuous waveform capnography. The 'no trace = wrong place' and PUMA guidelines both highlight the importance of assessing the capnography trace after intubation and while monitoring the patient. An analysis of closed claims relating to delayed detection of OI found that $ETCO_2$ monitoring issues was a factor in 73% of cases. In 27% of cases, the capnography device was not used at all due to a **lack of availability** of the device (Honardar et al., 2017). **Lack of immediate access** to important tools such as capnography, video laryngoscopy and other back-up equipment have led to cases of UOI (Chrimes et al., 2022; HDC, 2024). Not having adequate equipment prepared in advance has been reported as a barrier to timely airway management (Mann et al., 2023).

Organisational factors included the presence of team hierarchies hindering communication and teamwork, lack of training and inconsistent procurement of standardised equipment. **Hierarchies** negatively affect teamwork and can result in malignant politeness (Chrimes et al., 2022). This is an important consideration when 'crash teams' are brought together to manage emergency situations (Mann et al., 2023). **Lack of training** or awareness of the 'no trace, wrong place' concept was related to a fatal case of UOI (Flin et al., 2013). Prior to the publication of the PUMA guidelines, there was no universally agreed protocol for preventing UOI and so training and education has not been consistent. Organisational **procurement** decisions have a direct impact on the availability of equipment around the hospital, and with a lack of centralisation, can result in variations in equipment between hospital wards. This is a factor to consider as clinicians are often called to assist in emergencies where they may not have had sufficient orientation to the layout or equipment availability (Mann et al., 2023).

Task factors include **the difficulty and complexity** of inserting an artificial airway, which can be made worse under emergency situations. Determining the cause for an absent $ETCO_2$ trace involves many diagnostic tests with different equipment and can be considered a complicated process (Dob, 2022). The absence of standardised protocols can lead to inconsistencies in work practice (Cook et al., 2019). The **task sequence** can take on a variety of different versions depending on the situation at hand, relying on good situational awareness of the team to navigate the problem effectively. Skipping steps can occur during stressful situations and can be a symptom of cognitive biases causing lapses in situation awareness or premature closure (Nourallah & Levy, 2020).

Internal environmental factors that influence performance included **crowded and distracting working conditions.** A recurrent term used in coroners' reports is 'chaos' reflecting a stressful environment and deterioration in team function during the emergency (Chrimes et al., 2022). Critical moments in the task sequence require higher attention levels and even a moment of distraction can lead to incorrect tube placement (Holland et al., 1993).

New Zealand case studies

In 2024, two separate cases of UOI leading to death were reported in New Zealand (NZ) (HDC, 2024). This initiated the work to develop interventions addressing human factors issues raised in the adverse event analyses. These cases exhibit similar systemic risk factors highlighted in the literature review and called for the design of an intervention grounded in the local context of the NZ hospital system. Issues around the availability and awareness of back-up capnographs, as well as staff experience mix were highlighted in these cases. These themes address only a portion of the wider body of work related to prevention of UOI, with other themes relating to organisational factors such as training and teamwork being addressed in other bodies of work.

The two cases shared similarities which involved staff not knowing the existence of a back-up capnometer (used when the main bedside capnograph is unavailable or believed to be malfunctioning). In both cases, the bedside waveform capnograph recorded no trace when connected to the breathing circuit of the intubated patient. Clinical signs such as 'misting of the tube' were noted in both cases, which led the clinicians to believe that the ETT was correctly in the trachea, and the absence of a capnography trace was attributed to a faulty capnograph cable.

Confirmation bias and delayed use of bronchoscopy to confirm ETT location led to delayed detection of OI and unfortunately death of both patients. In both cases, it was later reported that the initial bedside capnography device was functioning as intended.

Capnography equipment can be fallible, and clinical staff should be made aware of the potential technical issues of older equipment. If there is doubt in the technical performance of this crucial equipment, there needs to be immediate access to a back-up device, such as a single use capnometer or portable capnograph. In both cases, staff members were unaware of the existence of these devices in the airway/resus trolley (i.e. in close proximity to the patient).

In one of these cases, it was reported that the lack of awareness of the back-up capnometer may have related to the poor labelling of the device's case. The capnometer sits in a case labelled EMMA provided by the manufacturer. This name is an abbreviation for EMergency Mainstream Analyser. Without specific orientation it is unlikely that any staff member would know that this a capnometer by purely reading the label. Additionally, the capnometer case was hidden from clear sight as it was at the back of an airway trolley.

In one case, registrars and anaesthetists were called in from different areas of the hospital to attend to the emergency case in the ED, where they had seldom worked. Unfamiliarity with the environment and equipment were associated with delays in recognising OI. In the other case, it was stated that the staff skill mix was below average on the shift when the UOI event took place. Additionally, this ward has had a high staff turnover rate which has resulted in a higher proportion of staff with lower than average airway management training.

Key interactions

The interactions between elements of the work system help to describe important aspects of the system that can have great impact on the overall system outcomes (Holden et al., 2013). The scope of this project addressed only the key findings relating to equipment (failure to confirm $ETCO_2$ trace and timely access to equipment). These themes resonated in the literature as well as the two NZ case studies and led to the formulation of design requirements for a product solution. These themes address only a portion of the wider body of work related to prevention of UOI, with other themes relating to organisational factors such as training and teamwork being addressed in other bodies of work. Table 1 below highlights the interactions that contribute to two issues: failure to confirm $ETCO_2$ trace and timely access to equipment. Each work system factor is uniquely coded.

Table 1: Interactions of work system relating to equipment

Themes	Person/team	Technology	Task	Organisation	Environment
Failure to confirm $ETCO_2$ trace	**P1:** Confirmation bias **P2:** knowledge deficit	**Te1:** Visibility of equipment **Te2:** Labelling of equipment	**Ta1:** Task sequence	**O1:** Interhospital staffing orientation **O2:** Lack of standard protocols	
Timely access to equipment	**P2:** Knowledge deficit	**Te3:** Equipment availability issues	**Ta2:** Equipment preparation **Ta3:** Time pressure	**O2:** Lack of standard protocol **O3:** Staff experience mix **O4:** Low staff training levels	**E1:** Room layout is not standardised

Design of cognitive aid

The scope of this stage of the project involved designing a cognitive aid to support airway equipment preparation to improve completeness and timeliness of this task. The items required for preparation were pre-determined by the project clinical lead and aligned with the globally accepted consensus guidelines for prevention of UOI (Chrimes et al., 2022). Design requirements of this tray were derived from the work system factors (Table 1) that were identified as being risk factors for poor airway management outcomes. Table 2 indicates which work system factors were addressed with each requirement, and the design specifications and success measures are stated alongside.

Table 2: Design requirements, specifications and success measures for tray development

Design Requirement	**Factors addressed**	**Specification**	**Measure**
1) Cognitive aid designed to assist nurses of all experience levels and familiarity with CVICU equipment	O1, O3, O4, Ta2	Equipment tray designed with shadow board concept to allow users to easily match equipment during equipment preparation	Equipment preparation speed and accuracy across range of staff experience levels.
2) Required equipment to be available in areas of hospital where intubations can occur	Te3, O2	Cognitive aid to be attached to new Glidescope Core, allowing for video laryngoscopy and bronchoscopy on single moveable stack.	Tray to be attached to Glidescope Core
3) Location of single use back up $ETCO_2$ monitor to be more obvious to all staff	P2, Te1, P2	Emergency capnograph located front and centre of tray so all staff have clear line of sight to it during intubation procedure	Staff able to quickly identify during usability testing
4) Back up capnograph can be clearly identified	Te2	Labelling to state capnograph function	Staff able to quickly identify during usability testing

Final design

A cognitive aid was designed by hospital product design engineers alongside a steering group of nurses. A tray identifying all necessary equipment through use of images and uniquely sized slots was fabricated using laser cutting technology (see Figure 2). The tray was fixed to a Glidescope Core which can perform both video laryngoscopy and bronchoscopy (Verathon, 2023). Having the tray affixed to the Glidescope Core ensures all equipment is brought to the bedside for all routine intubations, where previously it was observed to only be brought in when a difficult intubation was anticipated. A new label for the back-up capnograph was designed to clearly indicate its function (Figure 3).

Figure 2: Tray attached to Glidescope Core

Figure 3: New labelling of ETCO2 case

Formative usability testing

A pilot study was performed to gain an initial understanding of the effectiveness of this solution. The tray was tested with 10 participants to compare how this intervention impacts clinical workflow, relative to the existing workflow. Nurses are predominantly tasked with equipment preparation and therefore only nurses were recruited for this testing. A mixed-methodology approach involving a combination of observations, interviews, and questionnaires were used. Thematic analysis was used to review the insights and inform the iterative design process.

Procedure

Participants were tasked to prepare the equipment needed for intubation for a simulated scenario, using the existing airway trolley and the newly designed airway preparation tray which was fixed to the Glidescope Core. A total of 10 nurses of varying experience levels volunteered and were split into two study arms. Both groups consisted of two nurses with high experience (3+ years' experience in the ward) and three nurses with less experience (< 3 years). Group A performed the task using the current system first and then repeated the task on the new tray. Group B did this in reverse, using the new tray first. This was to counterbalance carryover effects within the testing.

At the end of the equipment preparation task, participants were asked to leave the room while three designated items were removed from their working space. The participant was brought back and then asked to identify the missing items. This identification activity was also performed after the participant switched systems where three different items were removed. The purpose of this task was to simulate the likely event of role allocation changes during an emergency response which may result in a nurse having to 'pick up' the equipment preparation task from a teammate. The ability to identify missing equipment is critical to the timely access of equipment when it is needed during the procedure.

Analysis

As this is an observational pilot study with a small sample size, the analysis was descriptive. The completeness of the preparation task was calculated as a percentage of the total number of items needed for intubation. The time taken to complete the preparation task was recorded, as was the time taken to identify the missing items.

Results

In both groups, an increase in percentage of completeness of equipment preparation was observed when using the new tray compared to the current system. On average, group A improved their completeness percentage from 62 % to 99 % (***$p<0.01$***). Group B improved from 83 % to 95 %. For both groups, the time taken to prepare the equipment using the new tray was higher, with group A taking on average 13 seconds more and group B taking 49 seconds more. For both groups, the time taken to identify missing items was lower when using the new tray than with the current system, with group A taking 7 seconds less and group B taking 25 seconds less (***$p<0.01$***).

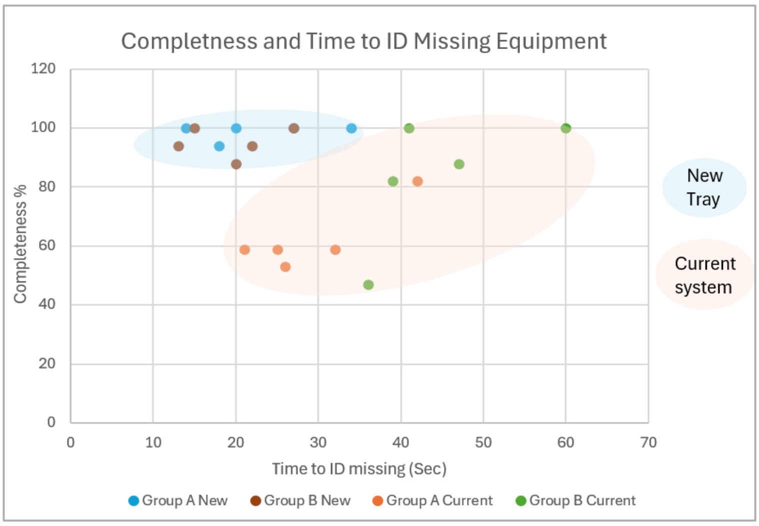


Figure 4: Graph of 'Time to ID missing equipment' vs 'Completeness of equipment preparation'

Table 3: Results of pilot testing

Measure	**Group**	**Current system (average)**	**New Tray (average)**	**Difference**	**T test (P value)**
Equipment prep time	A	110 s	123 s	13 s	0.35
	B	126 s	175 s	49 s	0.09
Completeness % (number of items)	A	62.4 % (11)	98.8 (17)	36.4 % (6)	0.001
	B	83.4 (14)	95.2 (16)	11.8 % (2)	0.10
ID missing equipment	A	29.2 s	22.6 s	-6.6 s	0.19
	B	44.6 se	19.4 s	-25.2 s	0.007

Discussion

In this pilot study, it was found that on average the completeness of equipment preparation as well as time taken to identify missing items was improved when using the new tray. Participants found that they relied much less on their memory and were able to easily prepare equipment by matching patterns. Different recall techniques were discussed by participants with many mentally simulating the intubation procedure to remember which items to prepare. Nurses commented that this technique is often flawed in high stress environments and equipment is easily forgotten. Addressing the issue of equipment availability, with the use of this tray correlating to an increased number of correct equipment prepared, it is more likely that all important equipment (including back-ups) will be available when needed.

All participants stated that it was much easier to identify missing items from the tray than it was with the current system as the technique involved spotting empty boxes, rather than relying on a mental checklist. This was reflected in the shorter time taken to identify the missing items using the new tray, in both groups. In emergency responses, role allocation can sometimes change for a variety of reasons. Having a clear template indicating what equipment has yet to be prepared is likely to reduce cognitive demand of an incoming clinician tasked with equipment preparation.

On average, the time taken to prepare the equipment using the new tray was higher than with the current system, which was to be expected as there was an increase in the total number of items prepared using the tray. It was also thought to be attributed to the lack of orientation participants had to the new tray with this being the first time all participants had seen the tray.

Nurses with more experience (3+ years) commented that they prefer the current system purely due to familiarity, however, were happy to change systems as they recognised it would be easier for those with less experience to perform the task effectively. This was confirmed by the nurses with less than 3 years' experience in the ward, with the majority stating they would "much prefer" the new tray and "felt more confident" they have prepared the necessary equipment with the tray. These findings indicate that this tray would suit nurses of varying experience levels.

Conclusion

A system level understanding of medical procedures helped to inform the design of a cognitive aid used to improve decision-making and performance of clinical staff in high stress environments. The intubation tray designed in this project places all routine intubation equipment on the same set of wheels as the video laryngoscope and bronchoscope. Given the central importance of a functioning capnograph, the tray also places a backup device front and centre of the airway management team. This tray standardises the equipment preparation process, considers clinician decision-making and may reduce the likelihood of future adverse events.

References

Baker, P. A., O'Sullivan, E. P., & Aziz, M. F. (2022). Unrecognised oesophageal intubation: Time for action. *British Journal of Anaesthesia*, *129*(6), 836–840. https://doi.org/10.1016/j.bja.2022.08.027

Carayon, P., Hundt, A. S., Karsh, B.-T., Gurses, A. P., Alvarado, C. J., Smith, M., & Brennan, P. F. (2006). Work system design for patient safety: The SEIPS model. *Quality & Safety in Health Care*, *15*(Suppl 1), i50. https://doi.org/10.1136/qshc.2005.015842

Chrimes, N., Higgs, A., Hagberg, C. A., Baker, P. A., Cooper, R. M., Greif, R., Kovacs, G., Law, J. A., Marshall, S. D., Myatra, S. N., O'Sullivan, E. P., Rosenblatt, W. H., Ross, C. H., Sakles, J. C., Sorbello, M., & Cook, T. M. (2022). Preventing unrecognised oesophageal intubation: A consensus guideline from the Project for Universal Management of Airways and international airway societies*. *Anaesthesia*, *77*(12), 1395–1415. https://doi.org/10.1111/anae.15817

Cook, T. M., Harrop-Griffiths, A. W., Whitaker, D. K., McNarry, A. F., Patel, A., & McGuire, B. (2019). The 'No Trace=Wrong Place' campaign. *British Journal of Anaesthesia*, *122*(4), e68–e69. https://doi.org/10.1016/j.bja.2019.01.008

Cook, T. M., Woodall, N., & Frerk, C. (2011). Major complications of airway management in the UK: Results of the Fourth National Audit Project of the Royal College of Anaesthetists and the Difficult Airway Society. Part 1: Anaesthesia†. *British Journal of Anaesthesia*, *106*(5), 617–631. https://doi.org/10.1093/bja/aer058

Dob. (2022). *Why does oesophageal intubation still go unrecognised?* https://doi.org/10.1111/anae.15692

Endlich, Y., Fox, T. P., Culwick, M. D., & Acott, C. J. (2024). Oesophageal intubations in anaesthetic practice across Australia and New Zealand: A webAIRS analysis of 109 incidents. *Anaesthesia and Intensive Care*, *52*(5), 302–313. Scopus. https://doi.org/10.1177/0310057X241244809

Flin, R., Fioratou, E., Frerk, C., Trotter, C., & Cook, T. M. (2013). Human factors in the development of complications of airway management: Preliminary evaluation of an interview tool. *Anaesthesia*, *68*(8), 817–825. https://doi.org/10.1111/anae.12253

Health and Disability Commissioner (HDC). (2024). *Te Whatu Ora breaches Code in care of man who died following incorrect intubation 21HDC02785*. Health and Disability Commissioner (HDC). https://www.hdc.org.nz/media/1d0bwyme/21hdc02785mediarelease.pdf

Holden, R. J., Carayon, P., Gurses, A. P., Hoonakker, P., Hundt, A. S., Ozok, A. A., & Rivera-Rodriguez, A. J. (2013). SEIPS 2.0: A human factors framework for studying and improving the work of healthcare professionals and patients. *Ergonomics*, *56*(11), 10.1080/00140139.2013.838643. https://doi.org/10.1080/00140139.2013.838643

Holland, Webb, R. K., & runciman. (1993). *Oesophageal Intubation: An Analysis of 2000 Incident Reports*. https://doi.org/10.1177/0310057X9302100519

Honardar, M. R., Posner, K. L., & Domino, K. B. (2017). Delayed Detection of Esophageal Intubation in Anesthesia Malpractice Claims: Brief Report of a Case Series. *Anesthesia & Analgesia*, *125*(6), 1948–1951. https://doi.org/10.1213/ANE.0000000000001795

Jafferji, D., Morris, R., & Levy, N. (2019). Reducing the risk of confirmation bias in unrecognised oesophageal int ubation. *British Journal of Anaesthesia*, *122*(4), e66–e68. Crossref. https://doi.org/10.1016/j.bja.2019.01.015

Kelly, F. E., & Cook, T. M. (2022). Unrecognised oesophageal intubation: Additional human factors and ergo nomics solutions. *Anaesthesia*, *77*(6), 718–719. Crossref. https://doi.org/10.1111/anae.15686

Mann, A., Higgs, A., & Cook, T. M. (2023). Preventing unrecognised oesophageal intubation. *British Journal of Hospital Medicine*, *84*(3), 1–9. https://doi.org/10.12968/hmed.2023.0007

NHS Improvement. (2021). *Never Events List 2018*. NHS Improvement. https://www.england.nhs.uk/wp-content/uploads/2020/11/2018-Never-Events-List-updated-February-2021.pdf

Nourallah, B., & Levy, N. (2020). Utilisation of 'verbalisation' to reduce the complications of tracheal intubation. *Anaesthesia*, *75*(4), 556–557. https://doi.org/10.1111/anae.15001

Russotto, V., Myatra, S. N., Laffey, J. G., Tassistro, E., Antolini, L., Bauer, P., Lascarrou, J. B., Szułdrzyński, K., Camporota, L., Pelosi, P., Sorbello, M., Higgs, A., Greif, R., Putensen, C., Agvald-Öhman, C., Chalkias, A., Bokums, K., Brewster, D., Rossi, E., … INTUBE Study Investigators. (2021). Intubation Practices and Adverse Peri-intubation Events in Critically Ill Patients From 29 Countries. *JAMA*, *325*(12), 1164–1172. https://doi.org/10.1001/jama.2021.1727

Sakles, J. C., Ross, C., & Kovacs, G. (2023). Preventing unrecognized esophageal intubation in the emergency department. *Journal of the American College of Emergency Physicians Open*, *4*(3), e12951. https://doi.org/10.1002/emp2.12951

Verathon. (2023). *Glidescope Core*. https://www.verathon.com/sites/default/files/2023-09/0900-5235_Rev-04_GS_TAS_BFlex2_SpectrumQC_WEB.pdf

Understanding work-as-planned and work-as-done in biomedical laboratories

Viji Vijayan

Biorisk Systems, Singapore

SUMMARY

An important concept in Safety II is distinction between Work-as-Imagined (WAI) and Work-as-Done (WAD). As in all industries, there are gaps between WAI and WAD in the biomedical laboratories, these concepts are poorly studied and understood. This study interviewed 15 biomedical lab workers to understand the gaps between WAI and WAD and how they were addressed.

KEYWORDS

Biomedical laboratories, workplace trade-offs, work-as-planned vs work-as-done.

Introduction

Gaps between work-as-imagined (WAI) and work-as-done (WAD) have been studied in many industries, but such concepts have not been applied to work in biomedical laboratories. WAI takes the form of Standard Operating Procedures (SOPs) and work instructions, developed with a mental image of how work should be done. WAD is the way in which work is really done. Safety in biomedical laboratories is managed through a Biorisk Management System (BMS) that addresses biosafety and biosecurity in an integrated manner. A fundamental purpose of BMS is to contain the risk posed by biological material where the work necessitates the presence of such material. Depending on the local work situations workers will not follow the written procedures exactly and will make adjustments based on experience and know-how. This study aims at understanding the gaps between WAI and WAD, the rationale of why the workers made the deviation and how the deviation is handled in biomedical laboratories.

Materials and Methods

Biocontainment is a combination of primary and secondary barriers, design of facility infrastructure and air handling units, facility practices and procedures, and safety equipment, including personal protective equipment (PPE). Biocontainment is described as biosafety levels (BSL) 1-4, with 1 being the lowest and 4 the highest level of containment. SOPs and work instructions are a key part of ensuring safety and productivity are maintained as part of the containment principles to ensure that workers and the environment are protected. BSL-2 is the most common type of biomedical laboratories. This study used convenience sampling method was used to invite 15 BSL-2 laboratory workers from different regions of the world to participate in the study. This study used the one-on-one semi-structured virtual interview method using key questions and topics (Table 1) to understand and explore the participants' opinions and experiences. This allowed the discovery of information which may be important to the participant but was not known to the interviewer prior to the interview thus allowing the interviewer to explore other dimensions of the research question.

Table 1: Interview questions

1. Demographic information 2. Are there safety SOPs/regulations in your lab? 3. Were you part of the team that established the safety SOP for your lab? 4. Do you deviate from safety SOPs/regulations i. Which is a common deviation? ii. What is the common reason for you to deviate from the safety SOP? 5. Do you know what Dynamic Mental Risk assessment is? *. i. Do you do Dynamic Mental Risk assessment 6. Do you think that the safety SOPs provides room for flexibility to accommodate diverse circumstances or varied conditions? 7. How to you communicate lessons learned or improvement in work instructions among your team: 8. How do you learn the lab work and the safety aspects
* Dynamic Mental risk assessment is the continuous process of identifying hazards, assessing risk, taking action to eliminate or reduce risk, in the rapidly changing work environment. The participants did not know the term and needed explanation by the interviewer.

Results

Standard Operating Procedures (SOP)

Clinical diagnostic laboratories provide diagnosis for patient care while research laboratories engage in research and development. Biomedical laboratories, both diagnostic and research have written step-by-step work instructions on how to perform the procedure for diagnosis or research. This ensures that the output is standardized and is of acceptable quality. In research laboratories reproducibility and replicability are very important to corroborate the results of the experiments. In clinical diagnostic laboratories, quality control ensures that the diagnostic results are reliable and consistent for patient care (Bayot, 2024; Diaba-Nuhoho, 2021).

As opposed to work instructions, Safety SOPs are developed by the organization, departments or laboratory heads as part of the biorisk management system. Safety SOPs address safety issues like: safe work practices, good microbiological practices and procedures, PPE requirements, waste management, establishing oversight committee(s) and many others. As this study is aimed at understanding the gaps between WAI and WAD solely for safety SOPs, the interviews did not focus on the gaps between WAI and WAD for the experimental/analytical work instructions. 14 of the 15 participants said that safety SOPs were not integrated into the work instructions. The one participant in whose laboratory safety SOPs and work instructions were integrated, said that the work instructions had references to the safety SOPs so that they would know at each step if any additional safety practices were required. Five participants said that the laboratory workers were consulted when developing the safety SOPs. The rest said that it was developed by the safety department or supervisor without consultation with the laboratory workers. The 15 participants each had very different understanding of how to integrate safety SOPs into the experimental/diagnostic work processes. One diagnostic laboratory had only a few written instructions for performing their work and no safety SOPs. Staff in this laboratory relied on word of mouth conveyed through the senior staff and the instructions were committed to memory.

Gaps between WAI and WAD

All participants did a mental risk assessment on the spot and made deviations to the SOP which resulted in gaps between WAI and WAD; these were never written down. All participants followed the SOP for certain steps like 1) using the biosafety cabinet (BSC), which is a crucial primary containment equipment, for work with live virus or samples containing a live virus; 2) Stock virus handling; 3) Virus Inventory Maintenance; 4) Waste Management; 5) Disposal of Sharps; and 6) Use of Personal Protective Equipment.

Among the participants interviewed there was a wide array of practices with one participant saying that very few rules were followed in their laboratories, and three saying that the workers followed the rules very strictly. The rest were somewhere in between, where the participants used mental risk assessment to make decisions. It is also worthy to note that the three laboratories where the workers followed the rules strictly had consulted the workers in preparing the SOP.

Examples of gaps between WAI and WAD

Use of PPE:

The core PPE requirements consist of laboratory coats, gloves and covered shoes. In a few laboratories additional PPE consisting of safety glasses and N95 or surgical masks were included in the SOP. All the participants wore the core PPE when working in the laboratories with no deviations. They wore a mask and safety glasses based on their individual mental assessment. The reason for this was cited as discomfort and they were not convinced that it was needed.

Sample transport:

The safety SOP contains an overall statement that all virus containing material must be transported within or between laboratories in double containment i.e. a primary container, usually a test tube to hold the biological agent and a leak-proof secondary container to contain the primary container. The SOP further stated that for transportation outside of the institution, double or triple packaging must be used based on the route of movement and the material being transported. Participants explained that within the laboratory they would carry infectious material in test tubes placed in an ice box which was not leak-proof. However, no one omitted the ice box because they wanted to keep the temperature of the material stable for the success of their work. All participants also used the correct packaging if they transported the material outside the laboratory or institution.

Use of fume cupboard for working with volatile chemicals:

Participants had access to a fume hood (device with exhaust ventilation that is designed for working volatile chemicals) which was usually placed in a central location for all laboratories to access. All participants did a mental risk assessment to decide whether to use the fume hood as stated in the SOP, based on the chemical, the quantity and the distance they needed to travel with the chemical bottle.

Incident reporting:

Incident reporting had a wide range of practice. In three laboratories where workers were consulted in preparing the safety SOPs, worker feedback as well as incident reporting was encouraged. Two participants from these laboratories, who held supervisory positions, said they had a system that at least 10 incidents should be reported every month, and each one was addressed with a strictly no-blame culture. The remaining 12 participants said that they would report only serious incidents. Seven of them said that reporting resulted in blame, and they would not report. The one laboratory with no written SOP also did not have any system to report incidents.

Reasons for gaps between WAI and WAD:

The most common reason for deviating from the SOPs is work pressure and time constraint (Figure 1). Participants had to ensure that the experiment was not jeopardized, especially in time-sensitive experiments with precious samples and reagents. In addition, participants chose convenience and the path of least effort, and if this worked and nothing untoward happened, they were likely to do it again. As they gained more experience, they would also omit steps that they thought were unnecessary and were not commensurate with the risk. This indicates that they relied a lot on their own mental risk assessment.

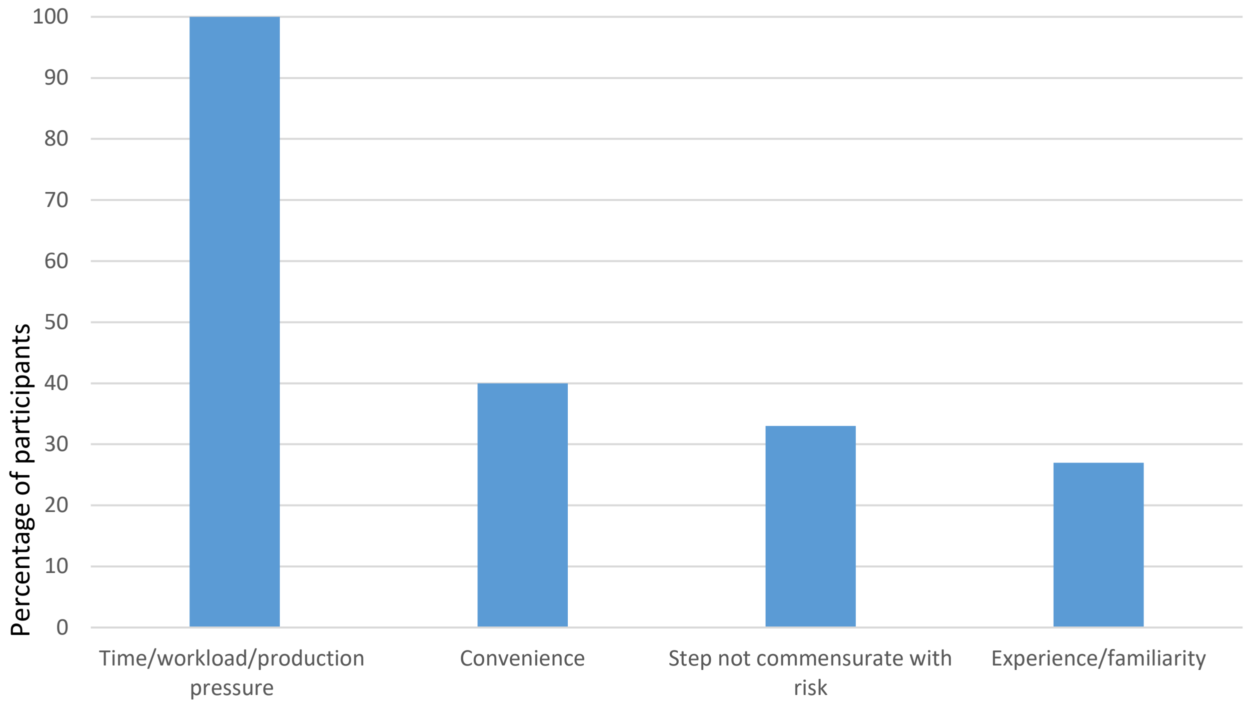


Figure 1: Reasons for gaps between WAI and WAD

Skills acquisition

Learning from their seniors was the most important method of skills acquisition. Only three said they learned from safety SOPs, and one from safety training. 47% of participants said they learned from information about accidents and incidents that happened to others and 20% said that they learned from others pointing out their mistakes. Regarding mistakes, 83% of participants said they learned from their mistakes and 13% said that if they got away with the mistake, they would do it again. When probed further about learning from incidents and accidents, the main reluctance appeared to be the assigning of blame that came with incident reporting and communication about incidents. One participant who is a supervisor said:

"When I was a junior the system in our institution is a blame system. I grew up in that kind of system where they blame you instead of correcting you or mitigating the problem so that's why in the course of my career here, I advocate the blame-free system, and we encourage incident reporting not to blame them but to mitigate the situation and to learn"

Discussion

This study shows that the laboratory workers do not always read the safety SOPs because they do not find it useful to do so. They said that the SOPs are repetitive and contain information that is not relevant to them. Based on published literature, this is not unique to biomedical laboratories (Eisner, 2022; Hammond Mobilio, 2022). Perhaps the SOPs are too prescriptive, for example, the use of a leak-proof secondary container to move infective samples a short distance within the laboratory and additional PPE that is not the normal practice in BSL2 laboratories. All the participants used mental risk assessment to assess the need for a leak-proof secondary container during transport of infectious material. They took into consideration the type of virus, the route and distance of the movement. This type of mental risk assessment is termed "Dynamic Risk Assessment" and has been used in industries with complex and dynamic work conditions (Jamshidi, 2018; Li, 2018; Paltrinieri, 2015).

All participants understood the need to use a BSC, which is the single most important primary equipment in biomedical laboratories. Class II A1 or A2 is the most used BSC, which not only protects the worker from aerosols generated when working with samples or cultures but also protects the samples and material from contamination (Jagtap, 2023; Pawar, 2021). One participant said that if they were working with an infective agent that was not aerosol transmitted but perhaps transmitted by a vector like mosquito, they would still use the BSC, but the reason would be to protect their samples from cross contamination rather than themselves. Such arguments and explanations of their work process show clearly that participants do understand the risks and use their on-the-spot mental assessment to decide whether to deviate from the written safety SOP.

Majority of the participants in this study said that they would not report or talk about any incidents or mistakes because they are associated with blame. Here, it can be useful to consider the idea of "just culture" where incident reporting and learning from incidents do not include assigning blame, unless the act was intentional with a malicious intent (van Baarle, 2022; Sieberichs, 2021; Amalia, 2019; Ale, 2020). The laboratories where the workers were consulted in developing the SOPs were also the ones where incident reporting without blame was encouraged. Learning from their mentors is the most critical knowledge transfer and skills acquisition method that is used. This highlights the need to pay attention to training senior workers to be good and effective mentors. Mentors need to teach both work and safety, remaining approachable and supportive (Hill, 2022; Roussakis, 2022).

Tresfon *et al* used the Functional Analysis Resonance Method to understand WAI and WAD in a hospital ward. The authors report that the use of this methodology gave the workers opportunities to reflect on the deviations and workarounds collectively as part of the study. One notable finding in this study was that the gaps between WAI and WAD were used to rewrite the SOPs, reduce the number of pages in the SOPs and change the narrative to a more friendly and helpful tone instead of a prescriptive tone of voice. Since the birth of the Safety II concept there have been many new ideas like safety differently, behaviour-based safety, safety culture among others. There is also heated debate on their usefulness and application. None of these are mutually exclusive and can be used in the same laboratory for different situations. In fact, there is no need to fixate on the nomenclature of these concepts as long as they are used to address safety and productivity at the same time. It is the workers who need to accept and deploy these concepts. We should think of the workers as a source of resilience instead of a problem to be controlled. Based on this very small and preliminary study, the first step is to encourage regular open discussion without assigning blame, include the workers in developing SOPs, accept that not all rules will be followed and study how to address the gap between WAI and WAD. More research is needed to understand WAI and WAD in biomedical laboratories, but in my extensive experience as a biorisk and human factors professional I am convinced that it is time to change and adopt newer concepts of safety to ensure that the gaps between WAI and WAD are addressed and minimised so that these deviations do not become normalised.

References

Ale, B., D. Hartford, and D. Slater, Resilience or Faith. 2020.

Amalia, D., Promoting Just Culture For Enhancing Safety Culture In Aerodrome Airside Operation. International Journal of Scientific & Technology Research, 2019. 8.

Bayot, M.L., et al., Clinical Laboratory, in StatPearls. 2024, StatPearls Publishing Copyright © 2024, StatPearls Publishing LLC.: Treasure Island (FL).

Diaba-Nuhoho, P. and M. Amponsah-Offeh, Reproducibility and research integrity: the role of scientists and institutions. BMC Research Notes, 2021. 14(1): p. 451.

Eisner, C., Your Standard Operating Procedure Program Failed (Now What?). 2022.

Hammond Mobilio, M., E. Paradis, and C.A. Moulton, "Some version, most of the time": The surgical safety checklist, patient safety, and the everyday experience of practice variation. Am J Surg, 2022. 223(6): p. 1105-1111.

Hill, S.E.M., et al., The Nature and Evolution of the Mentoring Relationship in Academic Health Centers. Journal of Clinical Psychology in Medical Settings, 2022. 29(3): p. 557-569.

Jagtap, G.A., et al., The Role of the Biosafety Cabinet in Preventing Infection in the Clinical Laboratory. Cureus, 2023. 15(12): p. e51309.

Jamshidi, A., et al., Dynamic risk assessment of complex systems using FCM. International Journal of Production Research, 2018. 56(3): p. 1070-1088.

Li, W., et al., Industrial non-routine operation process risk assessment using job safety analysis (JSA) and a revised Petri net. Process Safety and Environmental Protection, 2018. 117: p. 533-538.

Paltrinieri, N., F. Khan, and V. Cozzani, Coupling of advanced techniques for dynamic risk management. Journal of Risk Research, 2015. 18(7): p. 910-930.

Pawar, S.D., et al., Selection and application of biological safety cabinets in diagnostic and research laboratories with special emphasis on COVID-19. Review of Scientific Instruments, 2021. 92(8).

Roussakis, M., Mentorship in the Medical Laboratory Profession, in MedLab Scholar. 2022.

Sieberichs, S. and A. Kluge, How Just Culture and Personal Goals Moderate the Positive Relation between Commercial Pilots' Safety Citizenship Behavior and Voluntary Incident Reporting. Safety, 2021. 7(3): p. 59.

Tresfon, J., et al., Aligning work-as-imagined and work-as-done using FRAM on a hospital ward: a roadmap. BMJ Open Quality, 2022. 11(4).

van Baarle, E., et al., Fostering a just culture in healthcare organizations: experiences in practice. BMC Health Services Research, 2022. 22(1): p. 1035.

Working alone, saving lives: a focus on transfusion laboratory safety

Nicola Swarbrick, Debbi Poles & Shruthi Narayan

Serious Hazards of Transfusion (SHOT), UK

SUMMARY

Serious Hazards of Transfusion (SHOT) is the UK's independent, professionally led haemovigilance system, collecting and analysing anonymised information on adverse events and reactions in blood transfusion. Where risks and problems are identified, SHOT produces annual recommendations to improve patient safety. Transfusion laboratory errors reported to SHOT help highlight gaps in practices and policies paving the way for improvement actions to enhance safety. SHOT data shows that transfusion laboratory errors occur at a disproportionate rate when staff are lone working. A detailed analysis was undertaken of all laboratory errors reported to SHOT between 2020-2023 where staff working alone in laboratories was identified as a contributory factor (Narayan, 2023).

KEYWORDS

Lone working, human factors, blood transfusion

About SHOT

Since 1996, SHOT has been collecting and analysing anonymised information on adverse events, reactions and near misses occurring during the blood transfusion pathway, from donation through to hospital laboratory testing and administration in clinical areas. The SHOT team work closely with the MHRA as the regulatory body and other key transfusion stakeholders to enhance transfusion safety.

SHOT receives reports from all health organisations in the United Kingdom (UK) that are involved in the transfusion of blood and blood components. These events are submitted via a confidential electronic reporting system where a detailed questionnaire relating specifically to the reporting category is completed by the reporter, to ensure all relevant information is submitted including contributory factors and patient impact. Categories include serious adverse errors, serious adverse reactions and near miss events. Each case submitted is reviewed by SHOT incident specialists and working experts to identify trends and patterns, contributory factors, areas for improvement and implemented preventative measures. These findings are contained within the Annual SHOT Report (ASR) which is freely available online via the SHOT website and is widely shared within the transfusion community and to all relevant stakeholders.

Reporting to SHOT is professionally mandated, as SHOT is not a regulatory body. Through SHOT, organisations can monitor their haemovigilance reporting and benchmark themselves against organisations of similar size or blood usage to promote transparency, accountability and continuous improvement in blood transfusion practices. SHOT also receives reports of exceptional practice and innovative solutions to promote the learning both locally and nationally.

Haemovigilance reporting to SHOT is passive and reporting is limited to certain categories; not all transfusion related incidents are reportable to SHOT. It is also important to note that reporting levels varies between organisations and can be impacted on by local staffing issues and safety culture. SHOT is aware that there may be underreporting in certain categories, particularly in near miss events, so this data should be taken as indicative rather than complete. SHOT do not get involved in any local investigations and are dependent upon SHOT reporters submitting all relevant information via the reporting questionnaire or supplementary data. Reporters are encouraged to submit causal and contributory factors based on local incident investigations and applying human factors principles. It can be difficult to access complete data about the number of blood components transfused, the number of transfusions occurring out of hours and the number of laboratory tests performed within certain time frames.

Transfusion laboratory errors reported to SHOT help highlight gaps in practices and policies, paving the way for improvement actions to enhance safety. The SHOT questionnaire asks reporters to identify additional human factors which may have contributed to the error via the SHOT Human Factors Investigation Tool (HFIT). These questions cover the contributory factors regarding communication and culture, situational factors, organisational factors, external factors and lone working conditions. A detailed analysis was undertaken of all laboratory errors reported to SHOT between 2020-2023 where staff working alone in laboratories was identified as a contributory factor to understand common themes and inform improvement actions.

What is lone working?

Lone working can be defined as any situation in which someone works without close or direct supervision; without a colleague nearby or is out of sight or earshot of another colleague. Lone working can inherently pose a greater risk of errors as the worker is isolated, without colleagues available for advice or to assist in tasks. There are unique challenges lone workers face including high pressure decision making, increased responsibility, limited access to support through peers and senior staff, isolation and stress, confidence and competence issues, and communication and contact issues especially when dealing with emergencies in the laboratory where one staff member may cover several pathology disciplines during their lone working shift. Lone working is standard practice in UK laboratories, especially at night and during weekends. Prior to commencing lone working staff need to be trained and competency assessed, with risk assessment and support structures in place.

Blood transfusion laboratories – national regulations, standards, guidelines and recommendations

Detailed quality management system requirements for UK blood transfusion laboratories are outlined in the Blood Safety Quality Regulations (BSQR), governed by the Medicines and Healthcare products Regulatory Agency (MHRA), the competent authority overseeing the quality of the UK blood supply. MHRA has the responsibility to check that organisations adhere to these regulations via annual compliance reports and inspections.

UK blood transfusion laboratories must also adhere to quality requirements outlined by the United Kingdom Accreditation Service (UKAS) to gain accreditation to detailed standards through annual assessment visits.

Laboratory staff should follow all relevant guidelines and good practice guides outlined by the British Society for Haematology (BSH) in relation to transfusion practice to ensure provision of safe and effective blood components.

The UK Transfusion Laboratory Collaborative (UKTLC) is a collaborative group of transfusion laboratory representatives from across the UK, including SHOT, regulatory and accreditation bodies, external assessment schemes, and institutes and societies of biomedical science and blood transfusion. UKTLC produce standards for transfusion laboratories to develop staff knowledge and skills and improve transfusion safety. Compliance with UKTLC standards has been accepted by the MHRA, and UKAS as evidence to support their inspections for laboratories. UKTLC also conduct surveys to address gaps in practices, improve safety culture and promotion of best practice.

The Infected Blood Inquiry (IBI) report released in May 2024 highlighted the importance of laboratory safety and that laboratory teams can function optimally only if adequately staffed and resourced. The Inquiry reiterated the importance of staff receiving sufficient transfusion knowledge and training to reduce risks to patient safety by adherence to the regulations and standards provided by MHRA, UKAS and UKTLC. Reporting to SHOT and actioning SHOT recommendations was also outlined as essential in improving patient safety.

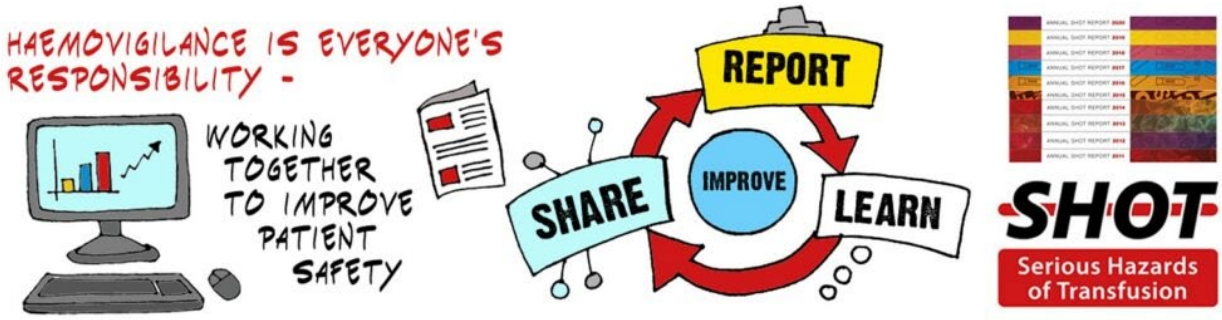


Figure 1. SHOT email signature highlighting the importance of haemovigilance in patient safety

Transfusion laboratory errors reported to SHOT between 2020-2023 with lone working identified as a contributory factor

There were a total of 1794 laboratory errors between 2020-2023, of which 540 (30.1%) occurred when the staff were lone working.

Most questions on the SHOT reporting questionnaire are non-mandatory, and therefore the denominator may vary due to non-responders.

There were 540 reports involving lone working laboratory staff, and these were analysed for contributory factors. Reports were assessed for the event time of errors, with 493 reporters submitting a response. Of these 263/493 (53.3%) errors occurred over a night shift between 8pm–8am, and 230/493 (46.7%) over a day shift between 8am–8pm. More specifically reporters were asked about the classification of working hours when lone working errors occurred and, of the 470 responses, 337/470 (71.7%) stated outside, and 133/470 (28.3%) within normal working hours.

Transfusion laboratory errors occurring during lone working resulted in:

- 171/540 (31.7%) blood transfusions which did not meet the patient's specific requirements
- 109/540 (20.2%) handling and storage issues of the blood components
- 86/540 (15.4%) laboratory-based patient identification errors
- 75/540 (13.9%) wrong component being transfused
- 61/540 (11.3%) transfusion delays which impacted on patient wellbeing
- 41/540 (7.6%) errors relating to anti-D Ig administration

These errors resulted in 2 patient deaths (1 possible, 1 probable) due to transfusion delays, and 11 cases of major morbidity due to development of anti-K in patients of childbearing potential due to laboratory errors. Of the errors 492/540 (91.1%) involved adult patients and 48/540 (8.9%) paediatric patients.

Many of these errors were compounded by lone working situations such single point of failure without a second member of staff to check or confirm correct procedure, gaps in knowledge, tiredness, difficulties in communication, dealing with emergency situations, multitasking and managing multiple laboratory areas as a single member of staff.

Contributory factors to errors

There were multiple contributory factors identified which compounded errors occurring during lone working. Most questions on the SHOT reporting questionnaire are non-mandatory, and therefore the denominator may vary due to non-responders.

Staff knowledge and training

In 350/540 cases reporters were asked if the member of staff had been competency assessed for the task where the error occurred. This question did not apply to all SHOT reporting categories until 2023. Of the 318/350 responses, 307/318 (96.5%) stated staff were competency assessed for the task, yet the error still occurred. In 370/540 cases reporters responded to the question 'were there issues or gaps with staff skill or knowledge', of which 128/370 stated yes there were gaps. Some reported cases mentioned that staff received accelerated training to expedite staff onto shift patterns due to staff shortages.

Information technology

Reports were assessed and 326/540 (60.4%) errors were deemed to involve information technology (IT). Common themes reported included overriding IT alerts, staff not heeding information available in IT systems, staff not adding IT alerts where available, lack of functionality in IT systems, patients having multiple records on hospital electronic patient records, lack of interoperability between systems and unplanned IT downtime. Use of IT to alleviate pressures from lone working staff has long been recommended, but staff must be sufficiently trained and IT systems implemented correctly considering human factors.

Task and team function

Reporters were asked to assess further contributory human factors in lone working errors:

- 159 of the 372 (42.7%) responders stated that task features impacted on the likelihood of the error
- 72 of the 369 (19.5%) responders stated there were failures in team function in relation to leadership, supervision, and roles

These errors are similar in other transfusion incidents occurring during routine hours but are compounded in lone working situations.

Specific lone working factors

Additional contributory factors stated in the reports included staff being physically and mental tired, lack of knowledge and experience, insufficient training, multitasking and distractions, insufficient staffing levels, high workloads, pressured emergency situations and sequential night shift working. Poor communication, poor handover, and IT issues compounded these errors. Local working conditions and situational factors contributed to the errors and included lack of out of hours support, poor delegation of tasks and staff being unprepared to deal with uncommon and/or emergency situations and complex cases. An illustrative case resulted in the wrong ABO group red cells being issued and transfused to a transplant patient. The lone working laboratory staff was under increased pressure due to a backlog of work remaining from the day shift and was rushing to complete all work before the end of their shift. IT alerts were not heeded, and training was deemed to have been insufficient.

The 2024 UKTLC culture survey identified that laboratory staff felt that work pressures impacted on their personal and home life, including negative impacts on their mental and physical wellbeing. A poor safety culture impacted on staff's ability to confidently and competently do their day-to-day role. Some staff also felt that they had been pressured to present an inaccurate assessment of the severity of an incident or been discouraged from reporting both internally within the organisation or to external bodies such as SHOT or MHRA.

Figure 2. SHOT email signature to highlight human factors contributory with blood transfusion errors

Mitigating strategies

SHOT

As part of the 2023 Annual SHOT Report, analysis of the laboratory errors highlighted several challenging areas of concern including abbreviated and accelerated training of laboratory staff to expedite staff onto continue processing patterns, and lone working laboratory staff. As such SHOT made several laboratory recommendations in 2023:

- Patients should not die or suffer harm from avoidable delays in transfusion. Where transfusion needs are complex, laboratory staff should have access to and follow specialist advice to provide the most suitable component available. Hospital policies and processes must reflect this.
- Staff must have protected time for training and education to provide a safe service.
- Policies for lone working should be reviewed to identify when extra support or reallocation of tasks are required.
- A just and learning safety culture should be implemented to improve the safety of patients and staff members, and to ease the existing recruitment and retention pressures in the laboratory.

These recommendations reinforced the need for staff working alone to be sufficiently trained, competent and confident in their tasks, with access to and follow specialist advice when required. Reallocation of tasks from lone working periods will alleviate additional pressures, allowing lone working staff to focus on the necessary tasks. Having a just safety culture where staff are well supported and confident to bring forward concerns and issues will lead to improved practices and enhanced service provision.

UKTLC

In 2023, UKTLC produced standards for transfusion laboratories (Dowling 2023) covering four main areas:

- Staffing - including staffing levels, capacity plans to meet workload demands, staff with specialist transfusion knowledge
- Qualifications, knowledge and skills - including minimum qualification requirements for all staff grades, training and competency assessment, and required provision of resources available to provide staff with sufficient knowledge and skills
- Information technology – including blood transfusion analysers, laboratory information management systems, and electronic transfusion systems
- Just culture – including encouraging staff to report errors, near misses, suggested improvements, potential risks that may impact on patient safety, and report examples of excellence, whilst consideration of human factors and system thinking

The UKTLC standards are produced with associated resources such as an example capacity plan and gap analysis self-assessment tools for teams to identify gaps in local processes. UKTLC also outlines the requirement for multi-disciplinary scientific staff and senior transfusion staff who would be expected to work within blood transfusion to complete a minimum of 10 practical working days per annum in a hospital blood transfusion laboratory. These standards are regularly reviewed to meet current workforce needs and requirements.

MHRA

The BSQR regulations include organisations having adequate numbers of personnel with the necessary qualifications and experience, and that management ensure staff are provided with adequate resources for the BSQR to be implemented and the quality management system to be maintained. MHRA state it is essential that laboratories have a capacity plan in place to ensure there is sufficient staffing to cover the workload, including out-of-hours working, and should be reviewed and updated if appropriate. This capacity plan should be available to suit all situations so business continuity can be preserved. Where gaps are identified senior management should take action to ensure sufficient resources will be made available.

Infected Blood Inquiry

The light of all the evidence supplied, the IBI made several recommendations relating to transfusion laboratory safety which included a laboratory and clinical training and education review including undergraduate and postgraduate to ensure they are adequately trained in transfusion, transfusion laboratories must be staff and resources adequately to meet the requirements of their functions, that all NHS organisations have a mechanism for reporting to SHOT and actioning their recommendations, and that a framework be established for recording outcomes for recipients of blood components.

Addressing workforce challenges

In England, based on the foundations of the Better Blood Transfusion initiatives, the 5-year Transfusion 2024 plan was to develop and promote the following key priorities - appropriate blood use, information technology, laboratory safety and research and innovation. The transfusion laboratory safety actions included the review of scientific training pathways and programmes to strengthen transfusion content and the provision of training resources to all laboratory staff to ensure an appropriately skilled and knowledgeable workforce. This has been realised in the format of the Transfusion Training Hub hosted via the National Blood Transfusion Committee website. This has been specifically created to support education and training for all healthcare professionals working within blood transfusion, providing a wide variety of learning, resources and signposts, each based on the levels of learning required from foundation and intermediate to in-depth and expert level.

Transfusion 2024 project is currently undertaking a review of all Institute of Biomedical Science (IBMS) accredited BSc Biomedical Science undergraduate degrees for blood transfusion course content. A working group of working transfusion experts and representatives from 12 universities has recently been established with the aim to produce standardised content to satisfy both the Quality Assurance Agency (QAA) standards and the IBI recommendations, which universities can access as a support tool to improve the content and delivery of transfusion education

The recent release of the Government's 2025 Mandate to NHS England has set out the vision for NHS reform and has placed biomedical scientists at the heart of these developments through improved efficiency and capacity of diagnostic services. In order to meet these challenges IBMS has called for:

- Expansion of training pathways and career progression opportunities to address workforce issues
- Increased investment in pathology networks and community diagnostic centres
- Greater recognition of biomedical scientists' expertise in strategic NHS workforce planning
- Clear career progressions pathways for biomedical scientists to maximise utilisation of skills and knowledge

The IBMS also outline the importance of widespread interoperability between digital systems to improve patient safety. Long-term investment in skills, technology and capacity rather than a focus on financial savings has been outlined as being key to the Government's plans.

Conclusion

Hospital transfusion laboratories are required to provide a 24/7 diagnostic and blood transfusion service to meet the needs of the hospital and its patients. Service provision requires laboratory staff to work out of routine hours, often lone working. There must be protective measures in place for lone working staff including risk assessments of tasks, regular reviews with feedback loops, availability out of hours advice or support, clearly defined standard operating policies and sufficient training and competency to equip staff to work alone. Policies for lone working should be reviewed to identify when extra support or reallocation of tasks are required.

There should be committment by teams to consider the impact of lone working on staff wellbeing, and the influence this may have on physical and mental health and an individual's home life, and mitigating actions taken where this is identified. Lone working staff require adequate rest breaks, and this must be factored into the laboratory's capacity plan. The capacity plan should demonstrated that the staffing levels are sufficient to cover the workload in all situations so business continuity is preserved, and reviewed and updated when required.

Harnessing the implementation of IT, and maximising of interoperability between IT systems will significantly improve patient safety. IT must be implemented correctly with consideration of human factors and ergonomics, to prevent workarounds or short cuts. Staff must be sufficient trained to use these systems with detailed contingency plans in place for downtime.

Acknowledgements

SHOT would like to acknowledge the contribution from the MHRA haemovigilance team, UKTLC members, and Transfusion 2024 members to ensure content is current and reflective of practice.

References

Blood Safety and Quality Regulations 2005 https://www.legislation.gov.uk/uksi/2005/50/contents

Dowling *et al* (2023). UK Transfusion Laboratory Collaborative: Minimum standards for staff qualifications, training, competency, and the use of information technology in hospital transfusion laboratories https://onlinelibrary.wiley.com/doi/10.1111/tme.13029

Institute of Biomedical Science https://www.ibms.org/resources/news/ibms-response-nhs-englands-2025/

Infected Blood Inquiry https://www.infectedbloodinquiry.org.uk/sites/default/files/Volume_1.pdf

Narayan, S. et al., 2021 - 2024. The 2020 - 2023 Annual SHOT Report, Manchester: Serious Hazards of Transfusion (SHOT) Steering Group. https://www.shotuk.org/shot-reports/

National Blood Transfusion Committee (NBTC) Transfusion 2024 programme https://www.nationalbloodtransfusion.co.uk/transfusion-2024

National Blood Transfusion (NBTC) Transfusion 2024 Transfusion Laboratory Safety https://www.nationalbloodtransfusion.co.uk/transfusion-2024/deliverable/transfusion-laboratory-safety

National Blood Transfusion Committee (NBTC) Transfusion Training Hub https://nationalbloodtransfusion.co.uk/transfusion-training-hub

NHS England national priorities 2025-26 https://www.england.nhs.uk/long-read/2025-26-priorities-and-operational-planning-guidance/#our-national-priorities-for-2025-26

Quality Assurance Agency for Higher Education https://www.qaa.ac.uk/membership/membership-areas-of-work/quality-and-standards

Serious Hazards of Transfusion (SHOT) Human Factors and Ergonomics investigation tool (HFIT) https://www.shotuk.org/human-factors-and-ergonomics-hfe/

Serious Hazards of Transfusion (SHOT) reportable laboratory errors in 2023 https://www.shotuk.org/wp-content/uploads/myimages/15.-Laboratory-Errors-2023.pdf

United Kingdom Accreditation Service https://www.ukas.com/

United Kingdom Transfusion Laboratory Collaborative https://www.shotuk.org/transfusion-laboratory-collaborative-uktlc/

From data to decision: A case study on ergonomics in manufacturing automation

Teemu Suokko[1, 2] & Arto Reiman[3]

[1]MSK Group, Kauhava, Finland, [2]School of Medicine, University of Eastern Finland, [3]Industrial Engineering and Management, University of Oulu, Finland

SUMMARY

Manufacturing process development is often technology and business oriented. The development of human work in this context is often neglected to a certain extent. With the use of ergonomics tools and methods, it is possible to identify targets for development and justify development investments also from the human work improvement perspective. For that purpose, physical ergonomics data were collected through video observations. In addition, occupational health and safety (OHS) and productivity indicator data were collected. In a series of group interviews, participants representing management and designers discussed the strengths and weaknesses of these data when it comes to decision-making on manufacturing development and related investments. As an outcome, the company decided to invest in production automation in the welding unit to avoid ergonomic problems and to increase productivity.

KEYWORDS

Manufacturing, Physical ergonomics, Video observation

Introduction

Manufacturing companies constantly consider means to develop their production performance and efficiency. Often, in manufacturing, human-technology and human-system interfaces are not optimised, resulting in different hazards for health and safety and lowering employee productivity. In this context, the use of ergonomics expertise would be highly beneficial. In reality, however, ergonomics is too often leaning towards siloed needs arising from occupational health and safety, whilst too little attention is paid to human productivity. While manufacturing development is inherently business-driven, companies with a broader vision for their future are increasingly considering how to gain a competitive edge through various factors, including the development of human work. In the current literature on manufacturing development, these—often paradigmatic—sociotechnical transitions are often discussed under the concepts of Industry 4.0 and Industry 5.0 (e.g., Grosse et al., 2023; Reiman et al., 2021).

As a scientific discipline, ergonomics aims to understand, design and develop work from a socio-technical work systems perspective. Arising from the complexities related to systems thinking in general, the concept of a system can be approached from different perspectives in ergonomics. A traditional way to discuss systems in ergonomics is to divide them into microergonomics and macroergonomics (also mesoergonomics on some occasions). When simplified, microergonomics focuses on individual-level work systems, whereas macroergonomics focuses on broader system complexities and their design and management (e.g., Kleiner, 2008). Usually, success in macroergonomics development requires microergonomics analyses focusing on the workstation level (Hendrick, 2003). Such microergonomics workstation level knowledge to be processed in further macroergonomics development processes can be collected with various means depending on the need (e.g., Lowe et al., 2019).

Case study approach and main findings

In this practical case study, we focused on one welding unit in a major-sized manufacturing company in Finland. The company has many subsidiaries in Central Europe. We discuss how physical ergonomics data, collected through ergonomics video observations, alongside actual OHS and productivity data from 2018 to 2023, eventually persuaded the company decision-makers to make significant changes in the welding unit to avoid the growth of ergonomics problems and to increase the unit's productivity. We present the video observation method, tailored for the company's purposes, and discuss its strengths and weaknesses. For this purpose, we conducted group interviews (n=4) for management and white-collar workers (15 persons). Finally, we discuss the process of how the company decision-makers eventually decided to invest in automatisation firstly to avoid ergonomics problems continuing and secondly to increase productivity. To concretise the development work done in the welding unit, we present in Figure 1 (on the left) one practical illustration of how the grinding task was earlier performed and (on the right) the design solution to automatise grinding work.

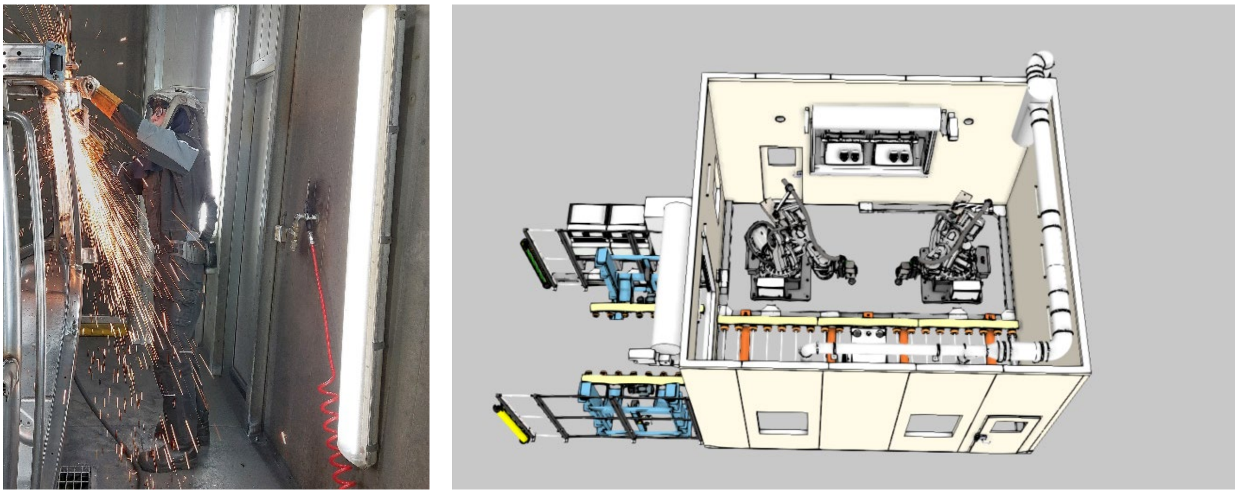


Figure 1: Grinding task with ergonomically hazardous working positions (left) and a design solution for a new automatised solution (right)

Discussion and key takeaways

Manufacturing companies need data on humans and technology to support their business development. This study shows how ergonomics data collected through video observations supplemented with actual OHS and productivity data can lead to investment decisions considerably higher than those being often in the ergonomics literature considered as “low-hanging fruits”. Through our interview approach, this study collected management and white-collar workers' insights into the usability of the ergonomics data collected and processed through the video observation process.

As the main author of this article works as the OHS manager in the case company, this study delivers a practical message from working life. A key takeaway is that ergonomics data is also of interest to the stakeholders if it can be tied to actual data on OHS and productivity. Thus, the data collected on ergonomics can be a good additional incentive for decision-makers to invest in manufacturing automation and technology.

References

Grosse, E. H., Sgarbossa, F., Berlin, C., & Neumann, W. P. (2023). Human-centric production and logistics system design and management: transitioning from Industry 4.0 to Industry 5.0. *International Journal of Production Research*, Vol. 61(22), pp. 7749–7759. https://doi.org/10.1080/00207543.2023.2246783

Hendrick, H. W. (2003). Determining the cost–benefits of ergonomics projects and factors that lead to their success. *Applied Ergonomics*, Vol. 34(5), pp. 419–427. https://doi.org/10.1016/S0003-6870(03)00062-0

Kleiner, B. M. (2008). Macroergonomics: Analysis and design of work systems. *Applied Ergonomics*, 37(1), 81–89. https://doi.org/10.1016/j.apergo.2005.07.006

Lowe, B.D., Dempsey, P.G., & Jones, E.M. (2019). Ergonomics assessment methods used by ergonomics professionals, *Applied Ergonomics*, Volume 81, 102882. https://doi.org/10.1016/j.apergo.2019.102882.

Reiman, A., Kaivo-oja, J., Parviainen, E., Takala, E-P., & Lauraeus, T. (2021). Human factors and ergonomics in manufacturing in the Industry 4.0 context - A scoping review. *Technology in Society*, 65, 101572. https://doi.org/10.1016/j.techsoc.2021.101572

Impact of a passive exoskeleton on human performance

Gonny Hoekstra[1]

[1]Technical University Delft , Netherlands

SUMMARY

This study evaluated the use of passive upper-limb exoskeletons in a bicycle assembly line to improve employee productivity and ergonomics. Results showed that exoskeletons improved posture and reduced back pain for tasks requiring work above shoulder height, but were less effective for tasks involving bending or varied postures. The research concluded that exoskeletons are beneficial for specific tasks but need careful selection based on the work environment, with further testing recommended for longer durations.

KEYWORDS

Exoskeleton, Physical Workload, Human Performance

Context

Within a funded project, which aims to investigate how robots and assistive technology can be used to provide inclusive and sustainable workplaces for all workers, one of the research areas focusses on physical enhancements (like wearable robots) in a real world context. In this research area a case study was conducted in a final assembly line of bicycles in a Bicycle company in Holland. The scope of the research in this case study was to integrate exoskeletons into manual production lines, improving physical ergonomics resulting in increased productivity (De Looze et al., 2016).

Actions

To achieve the aim of the research of the case study - evaluate the impact of passive industrial upper-limb exoskeletons on employees' productivity - data is collected, analysed and evaluated using different methods. To obtain a comprehensive description of the workplace within the assembly line, four evaluation methods were used, starting with a task analysis, followed by the Rapid Upper Limb Assessment (RULA) and the Key Indicator Method (KIM), an assessment of the level of automation and, finally, recording of external circumstances. For the description of the process and the operator within the assembly line various questionnaires and interviews were conducted.

Guided by the research plan, the bicycle company was visited several times to understand the assembly process, select the most suitable stations within the assembly line to test the exoskeleton and conduct a pilot test. The three stations chosen were those with the most tasks requiring work above shoulder height or a task with a longer duration of work above shoulder height.

The final research was conducted on several days, during which videos were made of the participants working on the stations, both with and without wearing the exoskeleton. Using these videos and the results of the task analysis, the postures were analysed using the RULA and the KIM. The RULA is a screening tool that assesses biomechanical and postural loading on the whole body with particular attention to the neck, trunk and upper limbs. The KIM is a survey method developed for use of assessing and designing physical workloads with respect to manual Lifting, Holding and Carrying of loads ≥ 3 kg. In addition to video recording, interviews were conducted and questionnaires administered both before and after wearing the exoskeleton. Finally, all the data was analysed.

Outcomes

The evaluation of the exoskeleton demonstrated a positive impact, particularly in terms of improved RULA scores for workplaces where employees need to work above shoulder height. In addition, there were positive experiences among participants wearing the exoskeleton ("My body is fixed", "I feel I work in a better posture" and "I experience less backpain during the task") and greater ease in specific tasks ("Lifting the motor is easier").

However, workplaces involving more varied tasks and postures, particularly those requiring bending, were found to be less suitable for exoskeleton use. Participants reported experiencing less freedom of movement.

Discussion

The study provides valuable insights into the impact of passive exoskeletons on the physical workload. While positive experiences were noted, it also highlighted the need for a nuanced approach, considering task complexity, variety and workplace conditions.

For the chosen workplaces within the bicycle company, using a passive exoskeleton is not the best solution due to the variety of tasks (not only above shoulder height, but also involving a lot of bending). Furthermore, other ergonomic issues within these assembly workplaces need to be addressed first.

Conclusion

In conclusion, we can affirm that the research setup and chosen methods worked well. However, the careful selection of the workplace (or task) for the application or introduction of a passive exoskeleton is essential. We recommend testing for a longer period in the field, specifically at workstations where employees have to work for extended periods at or above shoulder height. This would allow for a better evaluation of the impact of passive industrial upper-limb exoskeletons on employees' productivity.

Reference

De Looze, M. P., Krause, F., & O'Sullivan, L. W. (2016). The Potential and Acceptance of Exoskeletons in Industry. *Biosystems & Biorobotics*, 195–199. https://doi.org/10.1007/978-3-319-46532-6_32

On the use of ergonomic standards in Finnish manufacturing SMEs

Arto Reiman[1], Vesa Kauppinen[1], Maria Lindholm[2] & Susanna Mattila[2]

[1]Industrial Engineering and Management, University of Oulu, Finland, [2]Center for Safety Management and Engineering, CSME, Tampere University, Finland

SUMMARY

Ergonomics standards contain essential knowledge for the design of human work and human-technology interactions. When effectively implemented, this knowledge can significantly benefit companies by enhancing technological development, business performance and employee well-being. In this study, Finnish manufacturing small and medium-sized enterprises (SMEs) (n=16) were interviewed to assess the utilisation of ergonomics standards when designing human-technology interfaces during their endeavours to digitalise their manufacturing processes. The interviews indicate a general lack of awareness, and minimal application of ergonomic standards among the companies. This is a worrying observation, as it suggests that companies' design solutions may not always be based on current ergonomics knowledge but rather on alternative perspectives. To increase their competitiveness, it is recommended that these companies adopt a strategic approach to the utilisation of ergonomics standards. Future studies in diverse industrial settings and involving a broader set of companies are needed to confirm our findings and draw stronger conclusions.

KEYWORDS

Human work, Manufacturing, Standard

Introduction

European Union's (EU) Industry 5.0 vision (European Commission, n.d.) portrays a human-centric perspective for the future of working life. The vision is often considered a continuation of the earlier industrial revolutions that have revolutionised industrial work and human work alongside it. The most recent expression of these revolutions has been the so-called fourth industrial revolution - or Industry 4.0 - that is associated to broad workplace digitalisation. The corporate sector is still in many ways struggling with this overall ambition for this digitalisation, and for instance manufacturing SME's capabilities for successful new technology adoption have been seen immature in many ways (Ghobakhloo et al., 2022). In that regard, this qualitative study focuses on Finnish manufacturing SMEs and their organisational capabilities for human-centric design and management when seeking for competitive edge through manufacturing process development and related digital technological transitions.

There is considerable consensus in the research literature on how the central role of humans as users of digitalised new technologies has been overlooked. The kinds of physical, cognitive and psychosocial discomforts and loads or accident hazards related to the use of new digital technologies are still uncharted in many ways. Human productivity is also not always optimised when it comes to such new kinds of human-technology interaction situations (e.g. Alves et al., 2022; Grosse et al., 2023; Reiman et al., 2021). In that respect, it is good to remember that humans have not changed much in terms of their physical or cognitive characteristics, even though the intensity of work and technological demands have significantly increased during the last centuries (de Winter and Hancock, 2021).

Ergonomics standards contain up-to-date design knowledge to be used when designing human work. The standards provide a deeper understanding and interpretation of how to design human work and human-technology interactions effectively. Standards go beyond mere data by offering insights, principles, and guidelines that can be applied in practical contexts (Karwowski et al., 2021). The importance of using the knowledge contained in ergonomic standards seems to be raising increasing awareness with the EU decision makers. The recently launched EU regulation on machinery (2023/1230) points out the need for better ergonomics design activities and for instance, in a recently published Industry 5.0 road map for manufacturing (European Commission, 2024a) the use of ergonomics standards is highlighted in many ways. In another topical publication by the EU, designers' awareness of the existence of anthropometrics data—and also adequacies related to their presentation in current harmonised European standards—was critically discussed (European Commission, 2023). Paralleling with these EU level outcomes, there is evidence from the Finnish manufacturing companies—not depending on their company sizes, nor products manufactured—that the companies seem to lack certain organisational dynamic capabilities when it comes to design human work phases alongside the rapid technological development (Reiman et al., 2024a; 2024b).

Research approach

In an ongoing (2024-2027) Finnish research project, the aim is to increase understanding on Finnish manufacturing SMEs' organisational capabilities to meet present and upcoming challenges and opportunities that the digital technological development brings along from the human work design and management perspective. As one part of that project, companies are interviewed on their skills and knowledge to apply up-to-date ergonomics standards when they design their manufacturing processes and related new digital technology implementation (see Figure 1).

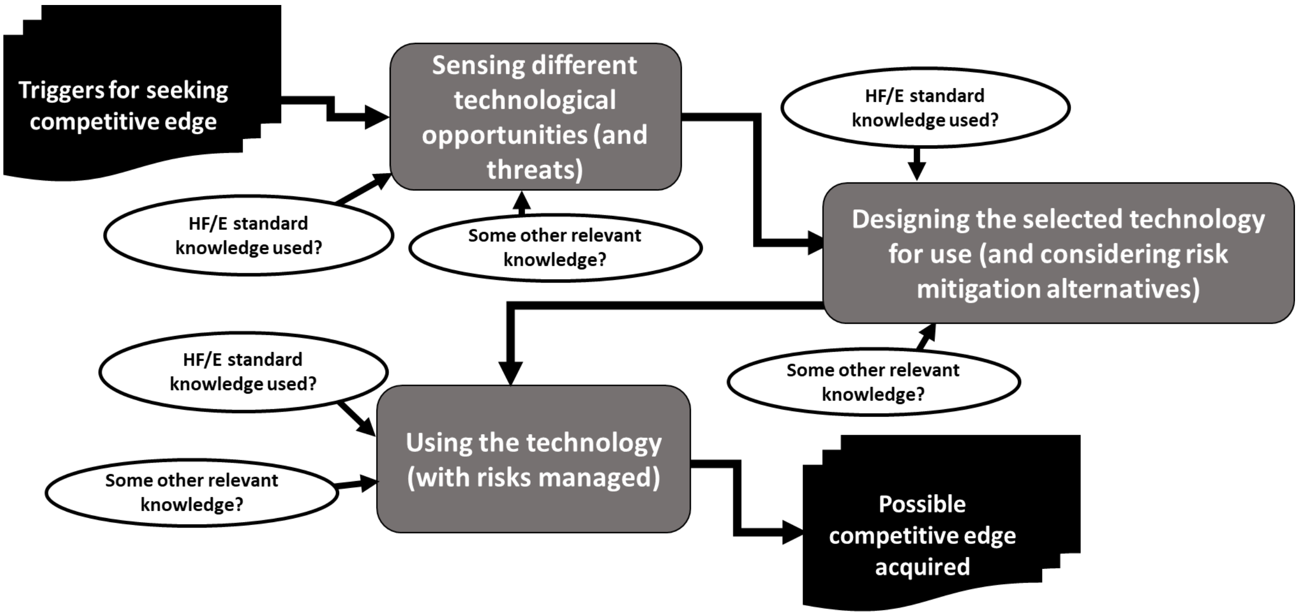


Figure 1: Research interests of this study

The interview guide with altogether 31 questions was created merging the existing knowledge on ergonomics in design (e.g., Karwowski et al., 2021; Reiman et al., 2021; 2024a), business development thinking (e.g., Teece, 2018) with insights arising from existing organisational level (maturity) standards on ergonomics (e.g., ISO 27500, 2017; ANSI HFE 400, 2021). To give an overall view of the interview guide, Table 1 presents the main themes of it and gives some examples of the questions addressing these themes.

Table 1: Interview themes and examples of specific questions

Theme	Examples of ergonomics related questions
Digital strategies and management	- … - What kinds of development needs have been identified for enhancing employees' skills and competences to utilise new digital technologies? - …
Identification of digital technologies	- … - How are the technologies' potential impacts on employees' physical and cognitive performance and well-being identified and considered? - …
Implementation of digital technologies	- … - How is it ensured in design that the use of a particular digital technology solution is ergonomic for employees? - …
Utilisation of digital technologies	- … - How are standards related to occupational safety and ergonomics monitored and applied in design? - …
Utilisation of data acquired	- … - What principles and processes are there in place for the collection, recording and evaluation of data related to human work? - …
Well-being at work	- … - What kind of health and safety threats related to digital technologies have been identified? - ...

The interviews started in November 2024. Altogether 16 companies have been interviewed. The interview process is continuing in the first half of 2025. In the interviews, company decision-makers, like managing directors and production managers, production designers, and occupational safety and health professionals, participate as informants.

Outcomes and key takeaways

Finland performs well in many areas of digitalisation (e.g. European Commission, 2024b), making it reasonable to examine the organisational capabilities Finnish manufacturing SMEs have in new digital technology transitions and how human work is considered in the transitions. Whilst the sample of interviews is small when it comes to the number of companies (n=16), the interviews were still uniform in a sense, that they revealed how immature the companies were when it comes to the broad use of ergonomic standards when designing the use of new digital technologies, and implementing them into practical use. In a few companies, the informants considered how some ergonomics standards had been used for machinery safety design purposes and when designing their end products. However, as an overall impression of the interviews, it can be concluded that they were not fully aware of the broad spectrum of the knowledge ergonomics standards would provide, nor did they identify that they would be in active use inside the company.

This research encourages companies to explore ways to advance human-centric industrial practices through the better use of ergonomics knowledge available in standards. This study is still ongoing, and further material will be collected in this respect. These initial findings seem uniform, suggesting that the use of ergonomics standards could be improved in manufacturing SMEs. Further research is still needed to draw stronger conclusions. Through this ongoing research process, the ultimate aim is to provide actionable insights into improving ergonomic design practices, supporting the transition to human-centric manufacturing in line with Industry 5.0 goals. Ergonomics standards have an essential role in this, as they contain scientifically proven information and knowledge on human limitations and capabilities, relevant to be understood when designing human-technology- and human-system-interfaces.

Acknowledgements

This article was partially written as part of the DigiCompetent project (203979), funded by the European Social Fund.

References

Alves, J., Lima, T. M., & Gaspar, P. D. (2022). The sociodemographic challenge in human-centred production systems – a systematic literature review. *Theoretical Issues in Ergonomics Science*, Vol. 25(1), pp. 44–66. https://doi.org/10.1080/1463922X.2022.2148178

ANSI/HFES 400 (2021). *Technical standard, human readiness level scale in the system development process*. Human Factors and Ergonomics Society, Washington, DC.

de Winter, J. C. F., & Hancock, P. A. (2021). Why human factors science is demonstrably necessary: historical and evolutionary foundations. *Ergonomics*, Vol. 64(9), pp. 1115–1131. https://doi.org/10.1080/00140139.2021.1905882

European Commission (n.d.). What is Industry 5.0?. Available at: https://research-and-innovation.ec.europa.eu/research-area/industrial-research-and-innovation/industry-50_en

European Commission (2023). Study on the inclusiveness of anthropometrics in European harmonised standards. Available at: https://op.europa.eu/en/publication-detail/-/publication/1712e683-b4ec-11ee-b164-01aa75ed71a1/language-en

European Commission (2024a). ERA industrial technologies roadmap on human-centric research and innovation for the manufacturing sector. Available at: https://data.europa.eu/doi/10.2777/0266

European Commission (2024b). Shaping Europe's digital future. Available at: https://digital-strategy.ec.europa.eu/en/policies/desi

Ghobakhloo, M., Iranmanesh, M., Vilkas, M., Grybauskas, A., & Amran, A. (2022), Drivers and barriers of Industry 4.0 technology adoption among manufacturing SMEs: a systematic review and transformation roadmap. *Journal of Manufacturing Technology Management*, Vol. 33(6), pp. 1029–1058. https://doi.org/10.1108/JMTM-12-2021-0505

Grosse, E. H., Sgarbossa, F., Berlin, C., & Neumann, W. P. (2023). Human-centric production and logistics system design and management: transitioning from Industry 4.0 to Industry 5.0. *International Journal of Production Research*, Vol. 61(22), pp. 7749–7759. https://doi.org/10.1080/00207543.2023.2246783

ISO 27500 (2017). *The human-centred organization. Rationale and general principles*. International Standardization Organization.

Karwowski, W., Szopa, A., Soares, M. M. (eds.) (2021). *Handbook of Standards and Guidelines in Human Factors and Ergonomics* (2nd ed.). Routledge.

Reiman, A., Kaivo-oja, J., Parviainen, E., Takala, E-P., & Lauraeus, T. (2021). Human factors and ergonomics in manufacturing in the industry 4.0 context – A scoping review. *Technology in Society,* Vol. 65, 101572. https://doi.org/10.1016/j.techsoc.2021.101572

Reiman, A., Kaivo-oja, J., Parviainen, E. Lauraéus, T., & Takala, E-P. (2024a). Human work in the shift to industry 4.0: A road map to the management of technological changes in manufacturing. *International Journal of Production Research*, Vol. 62(16), pp. 5613–5630. https://doi.org/10.1080/00207543.2023.2291814

Reiman, A., Takala, E-P., Parviainen, E., & Kaivo-oja, J. (2024b). Human work in strategic technology transitions in manufacturing. *International Journal of Human Factors and Ergonomics*, Vol. 11(3), pp. 266–288. https://doi.org/10.1504/IJHFE.2024.143256

Teece, D. J. (2018). Business models and dynamic capabilities. *Long Range Planning*, Vol. 51(1), pp. 40–49. https://doi.org/10.1016/j.lrp.2017.06.007

Deploying Usability Research Within Low-to-Middle-Income Countries

Alejandra Anderson Jimenez

Crux Product Design, UK

SUMMARY

This paper explores the human factors (HF) and usability research conducted during the early-stage development of a novel respiratory device designed to prevent the spread of respiratory viruses in a pandemic scenario. It highlights the application of methods within the disciplines of human factors, user-centred design (UCD) and ethnography in low-and middle-income countries (LMIC) to ensure the device design meets the diverse needs of a global population. The paper highlights the importance of an inclusive, adaptable approach to design and iterative testing to ensure effectiveness and accessibility across varied environments.

KEYWORDS

Human factors, user-centred design, ethnography, low-and-middle-income countries, global health, nasal respiratory device, COVID-19, inclusive design.

Introduction

In recent decades, the world has witnessed the emergence of several viral pandemics with devastating consequences on public health, economies, and societies, ranging from Human Immunodeficiency Virus (HIV) in 1981, to the H1N1 Influenza A Virus in 2009, the Ebola virus outbreak in 2013, and the most recent Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) in 2019 (WHO, 2022).

Whilst the characteristics of future viral infections cannot be known with certainty, the extensive investigations into the COVID-19 impacts provide a wealth of data with which we can seek to better prepare humanity for further pandemics. A snapshot of COVID-19 vaccination coverage revealed significant health inequalities between low and high-income countries (E. Mathiew et al.). Equitable global access to healthcare technologies is essential in controlling future pandemics.

The present usability work defined key requirements of a device suitable for the inhaled delivery of antiviral or prophylactic drugs to fight future pandemics.

A crucial differentiator in the development of such a device is the targeting of users in developing countries as well as developed, where limited financial resources and cultural differences make a high ease of use paramount.

Methodology

Human factors (HF) is the discipline that focuses on understanding the capabilities and limitations of human abilities and the application to such knowledge to the design of human-machine systems (Lewis, 2010). Within medical device design, HF involves a comprehensive understanding of the intended users, the use environment and the device user interface in the context of use, with the purpose of optimising the user interface to minimise use problems while ensuring a safe and effective use (FDA, 2016).

This is particularly complex when designing for a global population, which requires a broad perspective as the vast diversity in user characteristics and contexts of use must be considered to design a device suitable to use across the world. This wider scope requires the application of a novel approach, one that enables the application of methods to both capture real-world use behaviours and assess the usability of the designed solutions, enabling the rapid integration of knowledge into cyclical design iterations. To do this, we adopted a hybrid approach incorporating (1) user-centred design (UCD) and ethnography to capture real-life interactions with (2) HF methods to evaluate the user interactions with the device design. By leveraging the unique advantages of each field, integrating two approaches resulted in a robust methodology tailored to the design brief.

At Crux, our unique integration of design, engineering, and human factors expertise enables us to implement innovative approaches effectively, all within a single organization. Our multidisciplinary nature enables us to rapidly iterate and integrate user feedback to refine and optimise device designs. This flexibility allows us to tailor our approach to meet the specific project needs.

A summary of the methods deployed within this project are described below:

User-centred design (UCD). An iterative design process that prioritises the user needs, preferences, and limitations of end users throughout the design and development process to create highly usable and accessible products tailored to the end users' requirements (Interaction Design Foundation, 2016). UCD methods were deployed to understand global user needs, inform rapid iterative design cycles and guide the development or early-stage concepts through four phases: research, requirements, design and evaluation.

Ethnography. This qualitative research method was used to observe the behaviours and social dynamics of users in their natural environment (Lewis, 2010), exploring the influence of culture and social context on the adoption of a future respiratory device.

Human Factors. HF activities complemented the UCD broader approach, focusing on the study and evaluation of participant's interactions with the device's user interface, identifying potential use errors and informing the design optimisation. At this early stage, HF-specific activities conducted included a Task Analysis and the deployment of usability methods during user studies.

Throughout the four phases of the UCD approach, various activities were implemented, with user feedback continuously incorporated to inform design decisions and guide refinements across each of the following phases:

- *Research.* Primary and secondary research was conducted to investigate the key requirements of a device suitable for global use. Alongside usability work, investigations were deployed in the fields of life science (investigating the respiratory tract patterns), device design (investigating the effect of design choices in usability) and technology (uncovering development trends and novel technologies), (Quigg et al., 2022).
- *Requirements.* The research informed several technical, design and usability requirements intended to define the device operation as well as the user interactions with the device.
- *Design.* Usability and technical requirements informed the design of five early-stage device concepts. These were planned to be polarising to determine the suitability of different design features during the evaluation phase.
- *Evaluation.* Semi-functional prototypes were assessed by potential patients and healthcare professionals in real-use environments. Although label design was not in the scope of this phase of work, quick-reference guides (QRG) were provided to guide first-time use. Participants simulated use of the devices and provided subjective feedback and preference between the concepts and individual features.

As part of the research and evaluation phases, two user studies were conducted in Colombia (HFS001) and India (HFS002), nations chosen for their diverse healthcare systems, infrastructure, residential conditions, and cultural and geographical landscapes, each within a single nation. The sessions aimed to understand the use environment, current healthcare practices, user needs and insights surrounding the recent response to the COVID-19 pandemic, and the suitability of polarising prototype designs to support correct use in such environments.

The studies took the form of 60-minute, one-to-one semi-structured interviews, conducted in a variety of urban and rural locations within the participants' communities, including community centres and local healthcare facilities. Additional to these sessions, the study team conducted ethnographic observations to capture the real-use environments where the future respiratory device may be used. These included homes and workplaces.

Figure 1: User studies conducted in the rural areas of La Guajira, Colombia

The study recruited speakers of several languages, such as English, Spanish, Wayuunnaiki, Hindi and Bengali. Sessions were held in English or Spanish. If the participant was not able to communicate in either of these languages, an interpreter supported the translation of the interviews in real-time, facilitating communication between the moderator and participants. To ensure consistency, the Study Guide was available in English, Spanish, Hindi and Bengali and provided to the interpreter when needed. As the Study Guide was not available in Wayuunnaiki, the sessions were conducted in Spanish with the assistance of an interpreter.

A total of 34 interviews with urban and rural participants were conducted across the two studies, 18 in Colombia (13 potential end users, and 5 healthcare professionals) and 16 in India (8 potential end users, 6 healthcare professionals and 2 health facility administrators).

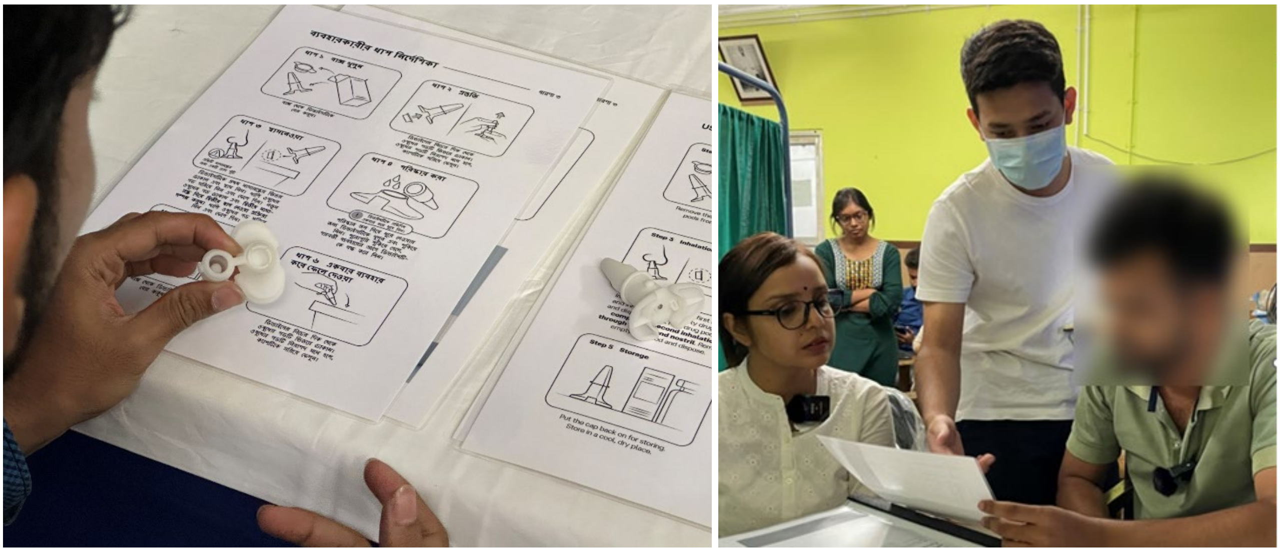


Figure 2: The study materials for HFS002 were translated in multiple languages

Main results

The first study in Colombia identified key challenges faced by both rural and urban users in accessing healthcare technologies and services. These key areas of opportunity informed the development of five polarising respiratory device prototypes, which were then assessed by users in the second usability study in India. The results from these studies were key in identifying factors that may promote correct device use in LMIC environments.

Environment: Facilities and storage

Interviews conducted in Colombia and India highlighted how families struggle to meet their basic needs, with limited access to clean water, refrigeration, and sanitation infrastructure, posing a significant barrier to proper device use. This includes in maintaining a controlled temperature, sanitising reusable devices between uses and disposing of them appropriately.

Furthermore, observation of the rural homes of the Indigenous Wayuu community (La Guajira, Colombia), constructed from basic materials like adobe and wood, evidenced the limited protection they offer from harsh environmental conditions. This underscores the need for devices that can withstand such climates, even when stored indoors.

Figure 3: Community kitchen within the Wayuu community, where a future device may be sanitised or stored

Cultural Acceptance: Engagement and suitability

The research evidenced how cultural and community dynamics could pose challenges to the adoption of a novel respiratory device. For example, the matriarchal culture of the Wayuu, with its strong adherence to customs and language, presents an additional barrier to embracing new health technologies, as the community tends to prioritise traditional medicine over Western treatments.

Furthermore, the study uncovered how negative perceptions surrounding the COVID-19 pandemic resulted in stigmatisation within communities. In Colombia, this was evidenced by the stigma faced by healthcare professionals during the vaccination campaigns. Whilst in India, these negative perceptions were evidenced by sick individuals who were rejected by their neighbours and often asked to leave their homes due to fear of disease transmission.

This underscores the importance of designing with thorough understanding of cultural behaviours, prioritising the design of an intuitive, easy-to-use device that minimises the need for assistance from healthcare professionals, and equally important, incorporates community-led strategies to address the community concerns and foster adoption.

Correct Use: Reduced instructional burden

During the prototype assessment in India, intuitive use emerged as a key factor contributing to correct device use. The findings revealed that single-use devices were perceived as more intuitive, as they required fewer steps for operation. Moreover, visual and/or auditory feedback was identified as a key design element, effectively communicating the device's status (e.g., ready to use or end of dose) and guiding users toward correct use.

Across all participants, challenges with reading and comprehending instructions were noted, stressing the importance of intuitive design to reduce cognitive load. Prototypes that featured a familiar form factor were perceived as easier to use, potentially minimising the need for instructions and assistance. This highlights the value of designing devices that align with user expectations and promote seamless, independent use.

Convenience of use: Driving user preference

The research uncovered that convenience was a key factor influencing device preference and adoption. The Wayuu community, for example, often travel long distances to access healthcare services, such as collecting prescriptions or attending appointments. In India, muslin-weaving workers provided invaluable insight into the work practices, highlighting how their busy routines – characterised by long commutes and limited storage space both on-the-go and at the workplace – made out-of-home use not suitable for their routines and needs during the day. These insights underscore the importance of designing a device that is compact and ease to store, enhancing convenience during transport and aligning better with user's routines.

Conclusion

The rise in pandemic events, such as COVID-19, underscores the importance of pandemic preparedness and the need for design solutions that help prevent the spread of diseases, ensuring that the global population is better equipped to handle future health crises.

This project has shown the importance of the application of novel approaches to capture real-world use behaviours, understand the influence of socio-cultural and environmental aspects on future device use, assess the usability of designed solutions and facilitate the rapid integration of insights into iterative design cycles.

At Crux, we combined methods from user-centred design, human factors and ethnography, creating a powerful approach that not only captured user interactions with the device user interface but also considered the broader context, including the healthcare system and the communities. The findings of this investigation evidenced that these considerations are fundamental for the successful design and adoption of a device intended for global use, particularly in LMIC, where key barriers to adopting respiratory treatments, such as cultural behaviours, a lack of trust in Western medicine, environmental challenges, and insufficient access to meet basic needs were revealed.

Furthermore, the evaluation of prototypes identified crucial design features that promote correct use in LMIC environments, including intuitive use, fewer steps for operation, designs that align with familiar products and increased convenience.

These insights stress the need for inclusive and adaptable design solutions that consider the cultural, environmental and socio-economic factors, while also addressing the needs of a global population.

Limitations and applicability

Although primary research was limited to two countries, in-field observations were consistent with the results of the broader secondary research, which covered a wider geographical area. Despite the two countries being geographically and culturally distinct, common socio-cultural behaviours, beliefs and challenges related to accessing healthcare and basic services were evident across both contexts. This suggests that the findings of this research may have applicability beyond the specific locations where the studies were conducted, and the design improvements derived from this work could have a far-reaching impact, benefiting a broader population in LMIC.

Acknowledgements

I would like to thank my colleagues at Crux Product Design Ltd for their support in this project. A special thanks to the Mechanical Engineering, Life Sciences and Design teams for their expertise in the development of the device design, incorporating usability insights, and generating prototypes and instructional materials that enabled the assessment during the study; to the Human Factors & Usability team for their collaboration in deploying secondary research which identified critical areas of investigation in subsequent fieldwork; and the study team (a multidisciplinary group including Project Management, Design, and Human Factors & Usability), for their dedication and logistical effort in enabling the deployment of in-field research across the two countries.

Additionally, I would like to acknowledge the local guides in both Colombia and India for their assistance in helping us connect with local communities and providing valuable insights into the local cultures and environments. Lastly, I extend my sincere thanks to the local communities who welcomed the study team, and shared invaluable insights into their daily lives and cultural practices.

References

World Health Organization. (2022). WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data. Available at: https://covid19.who.int/ (Accessed: 7 Sep. 2022).

Mathieu, E., et al. (2021). A global database of COVID-19 vaccinations. Nature Human Behaviour, 5(7), 947–953. https://doi.org/10.1038/s41562-021-01122-8.

Lewis, J. R. (2010). Human Factors Engineering. CRC Press eBooks, pp. 383–394. https://doi.org/10.1081/e-ese-120044161.
U.S. Food and Drug Administration (FDA). (2016). Applying Human Factors and Usability Engineering to Medical Devices. Available at: https://www.fda.gov/media/80481/download.
Interaction Design Foundation. (2016). What is User-Centered Design? The Interaction Design Foundation. Available at: https://www.interaction-design.org/literature/topics/user-centered-design.
Kirmayer, L. J. (2025). Ethnographic research in psychology: A cultural–ecosocial view. Qualitative Psychology, 12(1), 110–122. https://doi.org/10.1037/qup0000328.
Quigg, T., Harley, P., Pisa, A., Sefton-Collings, R., Fosker, N., Prescot, A., Hoare, J., Wade, B., and Muminov, R. (2022). Pandemic Preparedness, Exploring the needs of a respiratory device for the rapid global treatment of future viral infections. Crux Product Design.

Designing for a sustainable future: a user- and planet-centric approach to developing medical products

Pierre-Francois Gautier, James Ward & Paramesh Natarajan

Cambridge Consultants, UK

SUMMARY

There has been a trend towards single/limited-use disposable combination products to allow patients to self-administer treatments at home, in order to alleviate pressures on healthcare systems. These include inhalers, pre-filled syringes, auto-injectors and on-body injectors. In combination with this there has also been a tendency towards integrating smart technologies into some of these products to improve usability and patient treatment. These trends often conflict with the sustainability aspirations of manufacturers and pharmaceutical companies due to increase of waste, particularly electronic waste. It raises the question: Is it possible to develop a user- and planet-centric approach that supports the development of effective and sustainable medical devices and combination products, by marrying user-centred design and sustainability best practices?

KEYWORDS

Medical products, Sustainability, Life Cycle Assessment, User journey, Empathy map, Design thinking.

Introduction – The need to converge toward usable and sustainable devices

In this paper, we propose a user- and planet-centric approach and apply it to redesigning a hypothetical medical product, an automated insulin on-body injector, for the management of diabetes (Figure 1).

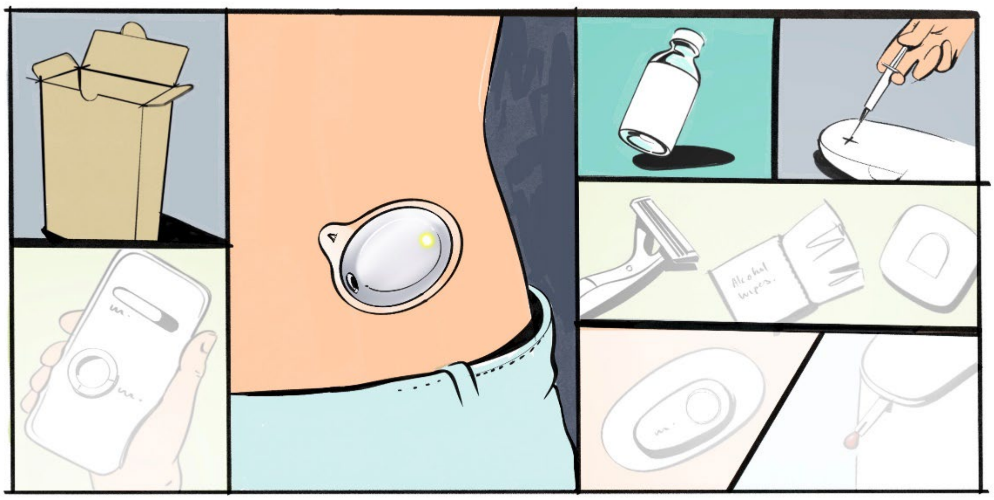


Figure 1: Sketch of typical on-body injector

Methodology

Combining approaches from a cross functional team

A multidisciplinary team encompassing skills from Mechatronics, Systems Engineering, Design Control, Sustainable Technologies, Industrial Design, and Human Factors Engineering, was formed with the goal of designing a more sustainable, user-friendly, safe, and effective automated insulin delivery system. The Human Factors workstream focused on several key use-related questions such as: what the main usability issues are with current devices; what their root causes and consequences are; when these issues arise during use; and how they impact user experience. The engineers employed standard Human Factors methodologies such as conducting a search for known use-related issues from MAUDE database (FDA, USA) and existing literature (e.g., patient forums, scientific articles), a task analysis within a Perception, Cognition, Action (PCA) framework, and creating personas to better understand user needs and behaviours:

- The known use issues search was tailored to extract what the main use errors were, what in the design of the device caused the issue, and what consequences it led to, with a particular focus on identifying situations that led to consuming additional resources, or prematurely disposing of and replacing the on-body injector.
- The task analysis was created to break down the use of the device and associate the use issues and their consequences found in the above activity to specific steps, to generate a high-level use error analysis. To gain a more comprehensive understanding of their struggle, the Perception, Cognitive, Action framework was applied to contribute to spotlighting potential areas of improvement.
- Personas were also created to provide a better insight on how users with different positions on sustainability may react to and adopt – or not – a new design that may create additional or different user tasks. The personas ranged from the environment friendly enthusiast to the busy parent, less concerned by the sustainability and more by the ease of use for their child to use the device safely and effectively, even if it means generating more waste.

In parallel, other team members aimed to understand the environmental impact of the on-body injector. A Life Cycle Assessment was performed to identify the key environmental impacts during manufacturing, packaging, distribution, use and disposal. By identifying the materials and their weights, and then manufacturing, distribution and disposal processes for the same on-body injector and co-products, its impact could then be calculated with the SimaPro (Amersfoort, the Netherlands) Life Cycle Assessment tool, using averaged impact data from the Ecoinvent (Zurich, Switzerland) database (Natarajan P, 2023).

The findings from both approaches were consolidated into a user journey and later an empathy map, presented in Figure 2, to visualise how the current design could negatively impact the user experience and the environment (Osorio S, 2024). These were discussed with the wider engineering team to input, comment, and question the content. Such diagrams have the advantage of providing easy to understand visual summaries, supporting collaboration with a wider team to review and configure the output to the project's focus needs. Additionally, a current and future architecture diagram was created. Architecture diagram is a visual representation which maps out the different parts and functions of the device. It was used for assigning key impacts from both usability and sustainability to certain functions, which made it clear where those functions should be allocated in the future architecture (presented in Figure 3).

Interaction between disciplines

Cambridge Consultants develops innovative technologies, products and services for a wide variety of clients in a range of different market sectors. To do this effectively, we are able to bring together a large range of engineers and scientists with diverse backgrounds. This collaboration was key for the project as it allowed having access to a multidisciplinary team to draw expertise and experience from many specialists, taking the form of many formal and informal exchanges where the team determined how they would work together. The initial questions addressed to the Human Factors team were related to the use steps, how frequently were they performed, and also ranking the usability of the marketed automated insulin delivery system in the scope of the analyses, what were the current use problems, and how often would they appear. Following this, the team had questions around the complexity and the duration of specific tasks. The findings were collated in the user journey using collaborative software tools.

Given the informality of these exchanges (messaging, workshops, remote or in-person meetings as needed) the advantage of having access to such multidisciplinary team was a major win-win situation, as it did not need an overly structured program of work. Such interactions may be difficult to reproduce with teams who do not work under the same entity.

Main findings

Identifying, qualifying, and quantifying issues

The findings from the different analyses identified elements of the current on-body injector design impacting both the usability and the sustainability:

- **Frequent replacement and refilling**:
  - Usability: The user must replace and refill the injector every three days using a vial and a single-use syringe, which requires precise handling and can lead to errors like underfilling (or not filling at all), spilling insulin or needle-stick injuries.
  - Sustainability: In the case of the insulin spillage and wastage, the Life Cycle Assessment revealed that the environmental impact was minimal in comparison to the other elements of the system.
- **Pairing issues**:
  - Usability: Following activation of the on-body injector, the user must pair it with the receiver (a smartphone-like device). If the pairing fails, e.g., from a poor user interface design, poor guidance on how to achieve the pairing, or software/electronic failure to connect, the user must start the entire process again (including filling). This means they would throw away the new, unused on-body injector, filled with insulin.
  - Sustainability: The Life Cycle Assessment showed that the Printed Circuit Board assembly was the most influential component across the range of environmental impact categories. The results revealed that device manufacture represents 75–95% of the environmental impact across all categories - global warming, ozone formation, land use, except one - marine eutrophication. Taking global warming as a representative example, two thirds of device impact are due to the printed circuit board and integrated circuit chips, which have a high demand for electricity during manufacture and extraction of raw materials. If pairing fails, these electronic components would be thrown away without ever being used.

- **Cannula deployment**:
    - Usability: Once paired, the user applies the on-body injector to the skin and deploys the cannula. If this step is missed or not completed, the whole device is to be discarded, and the process restarted. The known use issue search revealed that the cannula insertion was one of the most common sources of problems. Causes ranged from mechanical malfunctions at insertion to rather ineffective user feedback to confirm proper insertion (poor location and low contrast), which may have confused the users.
    - Sustainability: Similarly to the pairing issue, in the case where the user fails to insert the cannula, they may again throw the entire device away, and start the whole process over.
- **Wear time issues**:
    - Usability: The device is worn for up to three days, during which issues can arise that lead to premature disposal or failure to deliver therapy. These include the on-body injector coming loose, dislodging, or falling off from the infusion site, or the cannula bending, kinking, leaking, getting damaged or breaking off inside the skin from physical damage or weak adhesion.
    - Sustainability: These problems usually led to users disposing of the device prematurely and having to start the whole process over with a new device as the cannula cannot be re-inserted after removal from the skin.

The user journey framework was used and adapted to overlay the 'optimal' journey (or works as imagined, when the user does not have any issues with the device), and 'in real life' journey (or works as in reality, when the user faces problems with the device, as reported in the literature). The curves on the graph represent the emotional journey and any pain points that arise both from the usability and sustainability standpoints, presented in Figure 2. The crosspoints were designated as priorities for area for improvements.

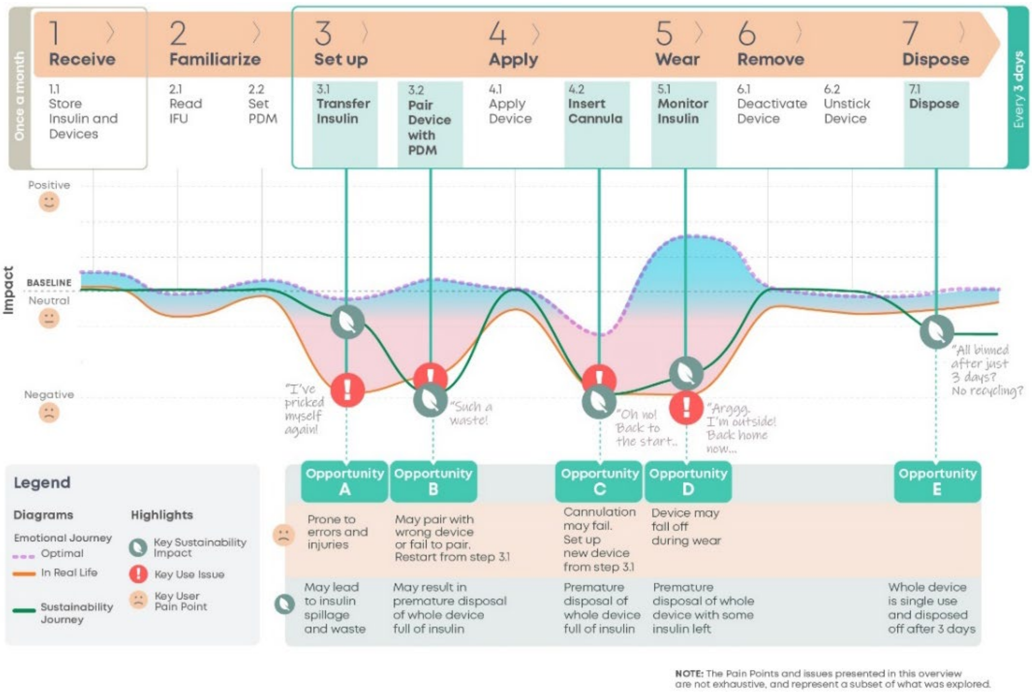


Figure 2: Empathy map overlaying pain points from the usability and sustainability standpoints.

Translating the findings into concepts

The team created a functional architecture for the current device, and using the outputs of the empathy map and the Life Cycle Assessment, the negative impacts were mapped onto each function. First, the current device was decomposed into high-level, solution-agnostic functions. Then, functions are categorised according to environmental damage, failure rate, cost, sterility and time consumption. By doing this, it became clear which functions belong on separate modules. The current and proposed future device architectures are presented in Figure 3.

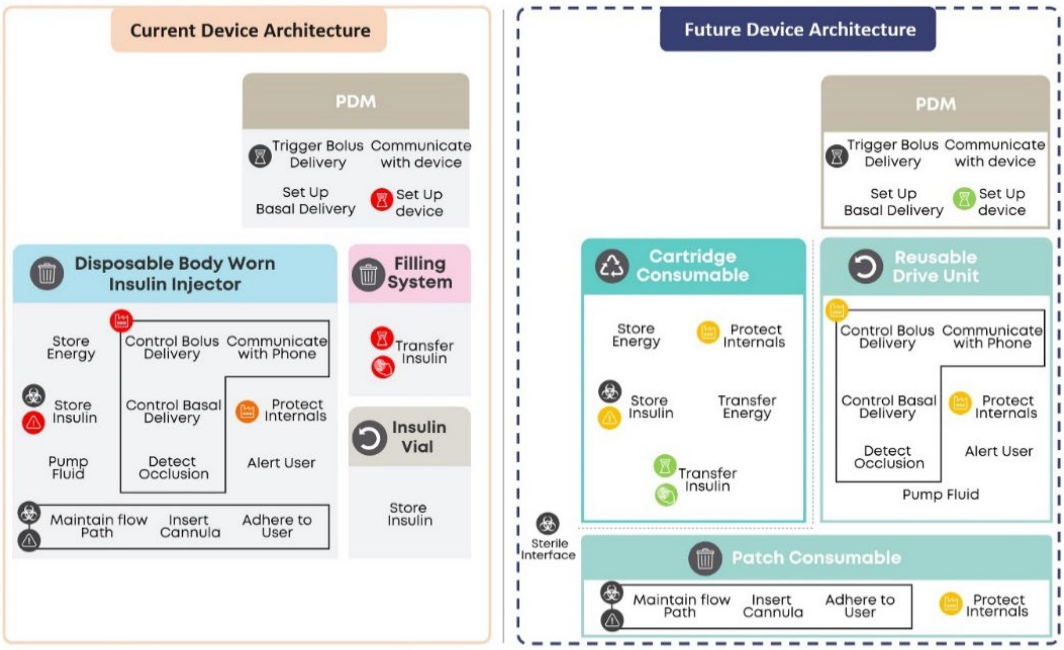


Figure 3: Device Architecture Redesign

Based on this new architecture definition, the team proposed a potential redesign for the on-body injector. The concept sketch is presented in Figure 4.

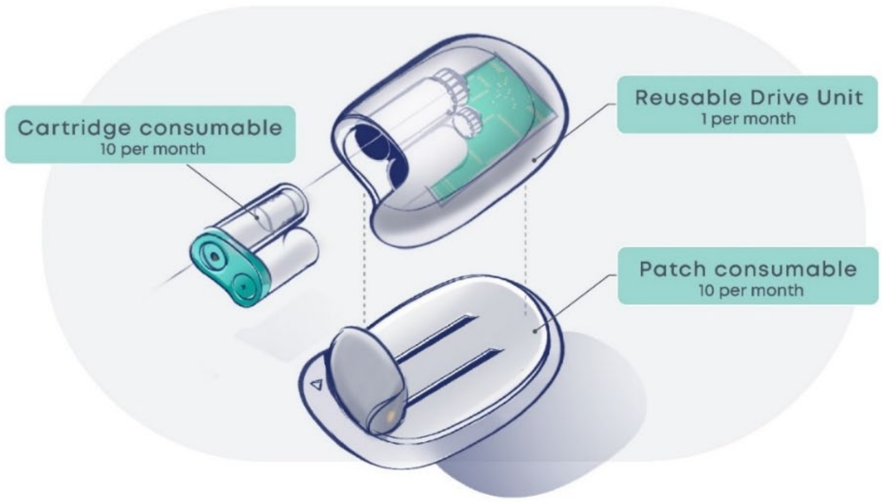


Figure 4: Concept sketch of a proposed future reusable on-body injector

The new design splits the device into three modular components to address the identified issues including insulin filling, pairing difficulties, cannula insertion and extended wear time:

- **Cartridge consumable**: Replaces the insulin vial and eliminates the insulin transfer user step. This designs out any problems with insulin filling. It also forgives situations where the on-body injector fails; the user does not waste anymore the insulin cartridges as it can be detached from the rest of the on-body injector, and reused. It also avoids certain practice with current on-body injector where users withdraw the remaining insulin with a syringe and refill the new device with this remaining insulin (not recommended by the instructions).
- **Patch consumable**: Includes adhesive, needle, and cannula components. In situations where the on-body injector either fails to stick or fails to insert, and keep the cannula in the skin, the user can replace this component without having to dispose of the whole device prematurely.
- **Reusable drive unit**: Contains the Printed Circuit Board, the most environmentally impactful component, and it can be reused over one month, rather than just three days. The key benefit is that it can forgive problems occurring with other separate components - needle and cannula insertion issues, and adhesion to skin problems - as the user now can unclip the components from the reusable drive unit.

Finally, the team put together a new user journey, presented in Figure 5, aiming to anticipate how the redesigned on-body injector is expected to perform, and how it may reduce the impact of certain pain points, both from usability and sustainability perspectives.

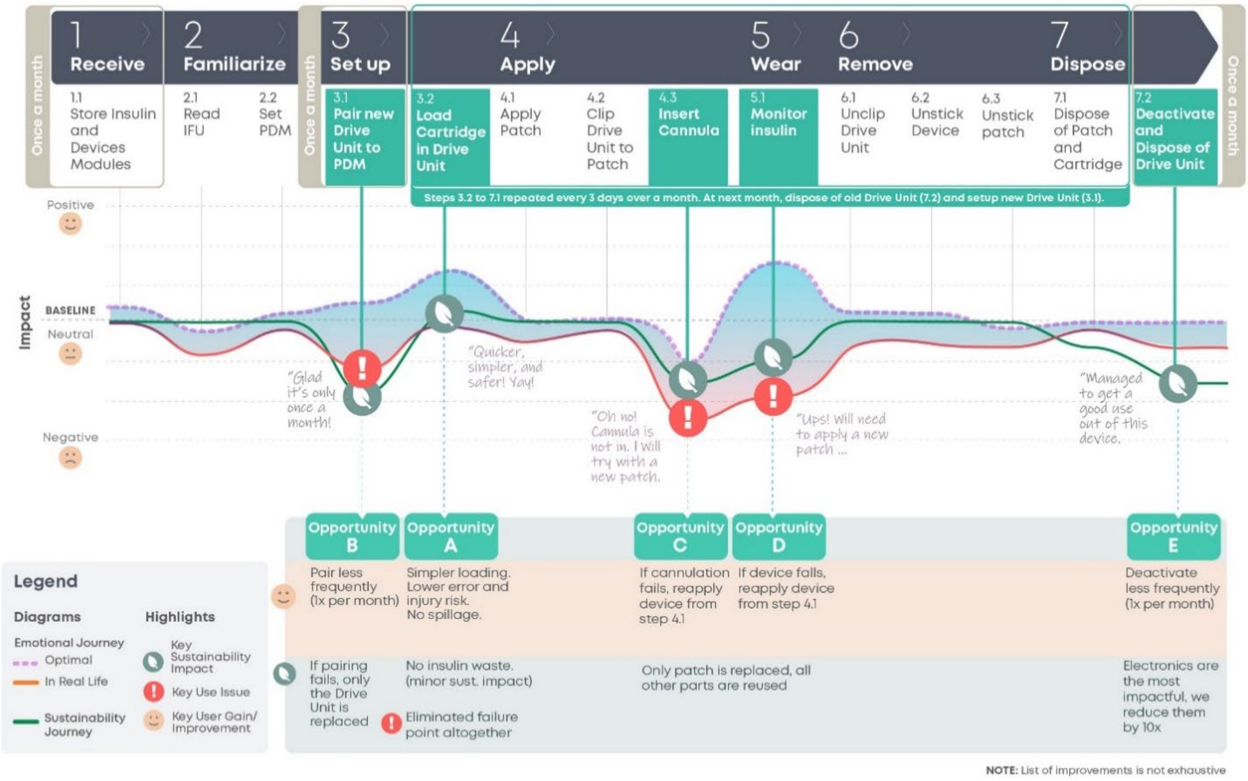


Figure 5: Future, anticipated user journey map

Key takeaways

At the beginning of this article, we raised the question of whether it is possible to develop a user- and planet-centric approach that supports the development of effective and sustainable medical devices and combination products, by marrying user-centred design and sustainability best practices. The answer is that by bringing together experts from different backgrounds, it is possible to craft innovative, data driven design that can be both less waste generative, and easier to use, by overlaying analyses from multiple angles.

Going forward, the following points highlights some of the key takeaways.

Opportunities for more holistic, data driven redesigns

This project demonstrates how integrating Human Factors methodologies into an initiative where a Life Cycle Assessment is carried out can lead to designs that are not only potentially more sustainable, but also more user-friendly. It presents an exciting opportunity for usability and sustainability experts, and manufacturers to contribute to a greener future for the next generation of medical products.

This also shows that combining user centred methodologies with knowledge from experts in sustainable technologies can bust myths around the generation of certain waste and their actual impact on the environment. Taking the example of the battery, the team initially thought about moving the battery to a reusable drive unit as it appeared to be the most sustainable option, to avoid disposing of the battery every three days. However, that move meant the user would need to recharge the device, necessitating a spare one that can be used whilst the other is being recharged. This would have created a higher usability burden compared to current devices. The Life Cycle Assessment showed that the environmental impact of the battery was less compared to the rest of the device, and therefore the battery has been designed as part of a disposable component of the proposed future reusable on-body injector. Without Life Cycle Assessment, this would have created an unnecessary pain point in the device's usability, with a minor gain sustainability wise.

Balancing ease of use with reusability

Creating reusable, more sustainable devices can also be challenging as it may require more effort from the user compared to the convenience of single/limited-use disposable devices. The creation of personas was useful to gauge the acceptability of the new design, and who may or may not embrace such changes: for example, it feels reasonable to assume that certain users with environmental awareness may be more inclined to adopt a device that may be less easy to use, but this may not be the case for other users who prioritize the comfort, and growing familiarity, of single/limited-use disposable devices.

Poor adoption is a threat to product's viability, and must be anticipated and addressed. In the proposed future reusable on-body injector concept, the following aspects will have to be closely monitored regarding their effect on the usability, safety and effectiveness:

- The user would need to assemble certain components before use, i.e., loading the cartridge consumable in the reusable drive unit, and clipping it to the adhesive patch. This was absent from current devices. While it removes the fiddly insulin filling step, it poses the problem of creating different, additional tasks, with new opportunities for use errors, failures, and generating new hazardous situations.
- Similarly, at the product's disposal, the user would now need to separate components rather than throwing away the whole device, i.e., removing the empty cartridge consumable from the reusable drive unit and unclipping the drive unit from the adhesive patch. Again, this was absent from the analysed device, as the user would discard it as a whole product.

Therefore, these tasks will have to be carefully crafted and evaluated through simulated use with end users, i.e., human factors formative studies. It should be as easy as possible for users to perform these tasks safely and effectively, including for those with dexterity, tactile and visual impairments from either diabetes or old age, or users experienced with the current design who may have difficulties adopting a new design because of negative knowledge transfer.

References

Osorio S, Mou S, Dean C, “Sustainable by Design: Developing Patient- and Planet-Centric Medical Devices”. ONdrugDelivery, Issue 159 (Apr/May 2024), pp 52–57.
Natarajan P, Bavar S, Dean C, “How Lifecycle Assessment Supports Insight-Driven Sustainable Design”. ONdrugDelivery, Issue 153 (Oct/Nov 2023), pp 16–19.

Human factors approach to platform device development

Finola Austin[1] & Miranda Newbery[2]

[1]Owen Mumford Ltd, [2]Inspired Usability

SUMMARY

Medical device manufacturers may enjoy some freedom to operate during platform device development. However, the absence of a specified drug means that there is a lack of predetermined limits and guidance with regards to various human factors that pertain to intended use, including user characteristics. This paper presents a best practice approach adopted by one manufacturer that aligns with the regulatory process and helps to anticipate the needs of a diverse group of potential end users. An inclusive approach to sampling both intended users and device variants is described alongside the resulting design decisions.

KEYWORDS

Medical device design

Background

Owen Mumford sought to introduce a 2-step autoinjector platform called Aidaptus® that has a similar mode of operation to several on market predicates. The autoinjector can house a range of different designs of 1ml or 2.25ml pre-filled syringes, with different fill volumes (0.3 to 2.0ml) and drug viscosities of ~20cP to deliver a drug formulation within a 3 to 10 seconds delivery time. As a platform product, it is intended that the product be suitable for use as a combination product by a wide range of intended users, but these are not yet defined and subject to onboarding of customers. Some of the initial design parameters are commercially driven. A human factors programme that aligns with international standards is planned and implemented to ensure that the needs of the unknown target audience are defined and supported throughout the design and development process. This also serves to manage use related risk and provide assurance to prospective customers about the product's suitability for their intended use case.

Human Factors Process

The human factors process dovetails with the manufacturer's design process and is aligned to medical device directives, international and best practice standards (see Figure 1).

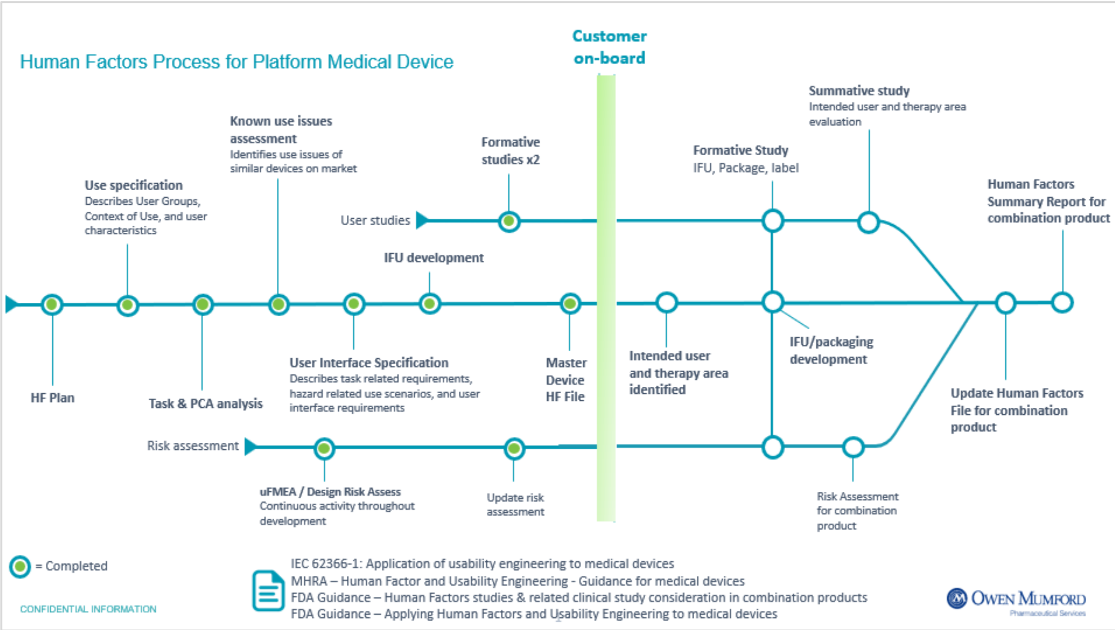


Figure 1: Human Factors Integration Process

Whilst the manufacturer assumes responsibility for the device, it will ultimately become part of a combination product, and the regulatory submission will fall to the pharma owner. As neither the regulatory pathway nor territory is known at that stage, the human factors process adopted a 'new product' approach, helping to mitigate the late identification of use related difficulties and risk for the intended user. The process uses a blend of activities from regulatory sources to ensure that there are no gaps in the human factors data that is generated; terms such as 'critical tasks' from the FDA guidance (FDA 2016; FDA 2018) are used, whilst the human factors file also includes a 'user interface specification' based on IEC 62366-1 (IEC 62366-1, 2020).

The Use Specification is drafted based on a series of assumptions about the most likely target audience and intended use scenarios. These are informed by:

- an understanding of general practice in the use of autoinjectors such as contexts of use and ambient conditions.
- a working knowledge of the autoinjector market, including existing and predicted therapy areas, and known end user needs related to features such as portability & size.

A broad use case was established to maximise the potential opportunity – with intended user age starting at 8[1], and a high-level summary of the range of intended user symptoms and associated impairments. Emergency use/lifesaving drug is excluded.

The preliminary analysis is designed to inform early design decisions and lay the foundation for any proposed user testing and is underpinned by use related risk analysis. A task analysis is carried out and used to identify which aspects of the user interface have potential to influence intended use at each use step. This not only supports the use related risk assessment, but it drives the scope and content of subsequent formative user testing. A set of generic use related hazards are defined with

[1] A child's psychomotor skills develop with age. A 2010 study of children injecting diabetes medication (Ekim, 2010) found that the success rate for self-injection developed rapidly over a child's age range, from as low as 16.7% in children aged 7-9 years up to 100% in those aged 13-18 years. Given the low success rate for younger children, and the high likelihood that parents would administer their injection, the lower age limit for the Juvenile group can be taken to be the midpoint age in the 2010 study, 8 years.

worst case assumptions for drug related aspects of the device such as drug contact and dosing errors. For completeness, all tasks are assigned 'critical' status, since there is no material advantage in not doing so; all device interactions would be user tested to ensure realistic use scenarios, and to generate data that can be available to prospective customers.

As the technical possibilities for the platform device is established, the breadth of design attributes and variants available to the user is confirmed. User interface specifications are again steered towards inclusivity, in the knowledge that they can be refined as more is understood about possible device constraints, and user capabilities and limitations.

IFU development

A generic instruction for use leaflet is drafted to support user testing. Since the primary focus is device usability at this stage, the size, format, layout and content were designed to get the best possible task outcomes. The manufacturer explored different images and text to communicate use steps effectively, guided by on market samples, and learning from concurrent in-house user testing.

The drafts did not include any level of customisation, branding nor drug related content that one might expect in a patient information leaflet.

Formative user testing

Sampling representative users: In the absence of a defined therapy area, the recruitment strategy used physical and sensory impairment, age, and role (e.g. patient, healthcare professional and lay caregiver) to select and screen participants resulting in 7 distinct user groups:

User groups

1. Healthcare Professionals

Patients/Lay Caregivers:

2. Children (aged 8-15)
3. Older adults (aged 66+)
4. Vision impaired
5. Hearing impaired
6. Musculoskeletal
7. Neurological

The sample was also screened for representation of handedness, gender, and level of injection naivety versus type of injection device experience to ensure a spread of relevant characteristics across each user group. The Cochin Hand Function Scale (Duruöz M.T. et al., 1996) was used to help select participants with a range of impairments that have potential to influence their interaction with the device interface. In addition to general demographic data, the study collected anthropometric measurements, and participants conducted dexterity tests to support the subsequent analysis of user task performance.

Experimental design

The platform device now had a range of attributes with variable potential to influence safe and effective use. The study was designed and executed by Inspired Usability to verify the efficacy of each solution to support design decision making; the range of study variants are presented in Figure 2, and the study design is summarised in Table 1.

Figure 2 - Aidaptus variants used in Formative Study 1

Table 1: Summary of device attributes and formative study 1 objectives

Device attribute	Attribute description	No. of variants	Study objective	Testing method
Syringe size	1ml & 2.25ml	2	The 1ml syringe is further away from the device outer surface than the 2.25ml – how does that affect visibility and user interaction?	Participants were assigned a 1 or 2.25ml variant for simulated use according to study stratification to ensure representation of each user group for each design.
Window size	The device variants include 5 different window sizes. The trial focused on 3 representative window sizes; Small 12mm x 8mm, Medium 25mm x 8mm, Large 39mm x 8mm	3	Window size would be driven by fill volume. Does this affect the user and their ability to inspect the drug and understand injection progress?	Participants were assigned different fill volumes for simulated use according to study stratification to ensure representation of each user group for each design. Drug inspection was also evaluated by comparing participants' ability to see different colour liquid in different prototypes
Window design	The device body concepts were developed – a moulded window, and a window that is created by an aperture in the label.	2	Does window design affect drug/plunger visibility and general handling?	Drug inspection was evaluated by comparing participants' ability to see different colour liquid in different prototypes

Device attribute	Attribute description	No. of variants	Study objective	Testing method
Injection speed	Injection speed is driven by drug volume and viscosity. Three nominal injection speeds of 3, 6, and 10 seconds were selected for evaluation	3	What, if any, effect on intended use and hold times?	Participants were assigned different fill volumes and injections speeds for simulated use according to study stratification to ensure representation of each user group for each design.
Audible feedback	Timing of audible feedback correlates with start and end of injection and therefore varied according to injection speed.	2	Establish the efficacy of audible feedback	Participants were asked specific questions related to their ability to hear and comprehend the audible feedback.
Cap design	A single concept was preferred. Several cap shapes were evaluated using a rig to simulate 25 and 30N pull force.	4	How easy or difficult is it to remove the range of cap variants. What is the impact of cap design on pull force ability.	A bespoke rig was created to evaluate different cap designs and different removal forces.
Plunger/ Window	Plunger inspection in different window designs and sizes	6	Establish whether participant is able to identify a used vs unused device	Participant inspects window in a range of devices that are used/not used to see if device is suitable for use
IFU	Instructions for use	1	Establish whether the device instructions are understood.	Observe use of IFU during simulated use. Further simulated use when directed to read IFU and comment on clarity. Knowledge based questions.

Two formative studies were conducted. The first study sample was controlled along with careful experimental design to ensure that a minimum of 5 participants in each user group e.g. older adult, or people with musculoskeletal impairment were exposed to each device attribute. The study sample was stratified and rotated across different configurations to ensure that each variant was evaluated by all user types represented in the sample. A similar approach was taken on a second study later in the development process as the device development was refined and updates to the user interface were made – notably the instructions for use. The sample size was significantly smaller but the group representation was maintained to optimise the study coverage. User testing was reduced where no further data was required.

Formative study testing was scheduled to ensure there was a sufficient level of fidelity in the prototypes to generate reliable and useful data. A generic instruction for use was created to support intended use. The study was conducted to the level described in international standards so that use related errors and issues could be quickly identified and mitigated where possible.

Analysis and Reporting

Formative study data was used to inform ongoing device development and continuing risk profiling and management.

The study outcomes showed with confidence that user task outcomes were not negatively affected by any single device design permutation. Experimental design coupled with study sampling means that the findings could be generalised to a wide range of intended therapy areas. None of the study findings lead to a requirement to change or limit the design of the device user interface and the range of variants.

Some changes were made to the instructions for use between studies – notably the steps related to the injection step and hold time. Whilst some improvements were achieved, it is known that users are not inclined to adhere to the correct hold time. No correlation between injection duration and hold time was identified. Some minor edits to instruction content have since been incorporated, but it is considered that there is no real advantage to the manufacturer to pursue this further.

Device Human Factors File

The human factors activities described above have been summarised in a device human factors file for Aidaptus®. It helps to establish its suitability for a wide range of applications and establishes functional limits. It provides assurance that the device design and development has been subject to a comprehensive human factors programme, in spite of the absence of a specified drug and intended use case.

The sampling technique has provided commercial benefits; typically pharma clients are keen to understand how their own target user group were represented in user studies during development. Clearly it has not always been possible to include every potential therapy area, but the inclusion of a wide range of sensory and physical impairments has meant that data is available for most applications. It is anticipated that the study variants have covered most foreseeable drug formulations, without evaluating extreme scenarios.

The author does not suggest that the work that has been completed negates the need for further human factors work once a pharma partner is on board. The specification of a drug will allow the use related risk assessment for the combination product to consider new information including but not limited to the specified user, characteristics of the drug, the injection, treatment regime, methods of distribution, training, context of use, and labeling for example.

Conclusion

The methods detailed above are aligned to international standards but go above and beyond that required of the device manufacturer to ensure the best possible outcomes for a wide and unspecified range of intended users. It serves to:

- identify the use related limits associated with what is technically possible in the device variants and
- minimise the gap between the device human factors file and the combination product human factors file.

Acknowledgements

The study was funded by Owen Mumford. The study was carried out by Inspired Usability with input from Owen Mumford.

References

Duruöz, M.T., Poivrade, S., Fermanian, J., Menkes, C.J., Amor, B., Dougados, M. and Ravaud, P., 1996. Development and validation of a rheumatoid hand functional disability scale that assesses functional handicap. *Journal of Rheumatology,* 23(7), pp. 1167-1172.

Ekim, A. & P., H., 2010. Insulin administration skills of children with type 1 diabetes. *14(2),* pp. 70-74.

FDA, 2016. *Applying human factors and usability engineering to medical devices.* Silver Spring, MD: U.S. Food and Drug Administration.

FDA, 2017. *Comparative analyses and related comparative use human factors studies for a drug-device combination product submitted in an ANDA.* Silver Spring, MD: U.S. Food and Drug Administration.

FDA, 2018. *Contents of a complete submission for threshold analyses and human factors submissions to drug and biologic applications.* Silver Spring, MD: U.S. Food and Drug Administration.

IEC, 2020. *IEC 62366-1: Medical devices Part 1: Application of usability engineering to medical devices — Amendment 1.* Geneva: International Electrotechnical Commission.

MHRA, 2021. *Human factors and usability engineering – guidance for medical devices including drug-device combination products.* London: Medicines and Healthcare products Regulatory Agency.

Realising market potential: HF and Design Thinking for novel ophthalmology patient interfaces

Phillips, T J & Gautier, P-F

Cambridge Consultants Ltd

SUMMARY

Breaking into a crowded space requires compelling differentiation. Our client wanted to enter the ophthalmic scanning market with a product which prioritises patient throughput to support the viability of optician businesses. We adopted a unique approach to design-thinking, prototyping and testing to explore novel patient interfaces to help achieve this challenging product requirement. The focus was not solely on enhancing ergonomics but moreover on supporting a robust business case - the surprising results may unlock significant commercial opportunity.

KEYWORDS

Medical devices, Ophthalmology, Commercial Opportunity, Design thinking, Formative Testing, Data analysis

Introduction

The demand for diagnostic technologies continues to grow as advances in medical science expand our ability to identify and treat disease. Within ophthalmology, this trend is particularly significant, as early detection has proven pivotal in managing a range of conditions, from glaucoma to age-related macular degeneration and diabetic retinopathy. Optical Coherence Tomography (OCT) has emerged as one of the cornerstone technologies in this field, offering high-resolution cross-sectional images of the retina that guide both diagnosis and treatment.

Despite the clinical value of OCT, improving access to this technology remains a pressing challenge. A wide variety of OCT systems have been developed, creating an increasingly saturated market where new entrants must offer meaningful differentiation to succeed. Recognising an opportunity, our client—a developer of premium multi-modal ophthalmic systems—approached us with a goal to enter the OCT market. Their vision centred on a device tailored to meet the needs of optician businesses by prioritising patient throughput and minimising attendant interaction, thus providing a competitive edge by supporting faster profitability.

This paper details the design and development process we undertook, emphasising the role of human factors engineering in shaping patient interfaces. By integrating design thinking, iterative prototyping, and innovating user testing, we explored novel interfaces to only enhance ergonomic comfort but also address the critical business case for optician practices. The insights gained offer a case study in how user-centred design can drive both clinical and commercial success.

Methodology

Medical product development follows a structured process to ensure devices are safe, effective, and meet user needs. It typically progresses from early exploration and concept generation, through formal development stages governed by regulatory requirements, and ultimately to summative validation. Each phase builds on the previous, with early insights shaping later development.

The presented work fits into the front-end, pre-regulated phase, driven by design thinking principles. At this early stage, the focus was on exploring and validating concepts rather than adhering to formal design controls. The overarching goal was to provide the client with a 'reason to believe' that their commercial objectives could be achieved through innovative design and engineering solutions.

In this project, the client sought to differentiate their OCT system by maximising patient throughput—a key commercial requirement for optician businesses. To address this challenge, a two-part program was conducted: (a) an ideation and rapid prototyping phase to explore potential solutions, and (b) an in-depth evaluation of lead concepts to assess their compatibility with other product requirements. Although exploratory, the process incorporated elements of conventional Human Factors practice, such as task analysis and risk-based evaluation, ensuring meaningful and rigorous assessment.

The project tested the hypothesis that redesigning the user interface could reduce patient contact time while maintaining required standards of eye fixation and stability. These findings provided early confidence that the design could meet both clinical and commercial goals, offering a strong foundation for future development phases.

Although this program was an early-stage design-thinking exploration, a number of conventional Human Factors (HF) methods were selectively applied to ensure meaningful insights while maintaining agility in concept development. By balancing conventional HF methodologies with rapid, exploratory testing, the program maintained creative flexibility while ensuring design decisions were grounded in usability and anthropometric data.

Task Analysis & Risk Exploration

A high-level task analysis was conducted for multiple market-leading OCT systems to map out typical user workflows. Rather than a detailed step-by-step breakdown, this analysis remained abstract, focusing on identifying key operator and patient interactions that impact efficiency. The goal was to pinpoint areas where design innovations could streamline workflow and improve throughput.

While a formal uFMECA process was not followed —given the exploratory nature of this work— informal risk discussions were held with the client. These sessions explored ‘pinch points’ and usability challenges that could lead to negative clinical and business outcomes, helping to shape design priorities without constraining creativity.

Ideation and Prototyping

In understanding the conventional workflows and processes, the design thinking methodologies moved to focus on novel patient interfaces. Conventional OCT systems typically rely on an adjustable chin-and-headrest assembly, which can be time-consuming to adjust and uncomfortable for patients. This presented an opportunity for innovation. A range of alternative interface concepts, including ‘goggle-style’ and ‘monocle-style’ designs, were hypothesised and prototyped to varying degrees of fidelity using a various methods & materials.

Early Usability & Anthropometric Evaluations

A number of small-scale usability studies were conducted as part of iterative, fail-fast design sprints. These tests helped to stress-test interface concepts, and progress to the down-selection of two lead candidates for focussed evaluation. These internal studies leant upon a diverse sample of employees and relatives (including elderly and child participants) to before committing to the more structured formative assessment described in this paper.

With lead concepts selected, a digital anthropometric analysis was undertaken to determine the appropriate dimensions for the lead concepts to accommodate the target population, and assess whether individuals with extreme facial morphologies could use the interfaces comfortably without interference with the internal optics system.

This analysis processed first identified the key facial dimensions which would impact design and determined the corresponding anthropometric values from the PeopleSize [OpenErgonomics Ltd.] database (utilising 1st to 99th percentile values). Next, a number of digital personas were modelled within CharacterCreator3D [Real Illusion Inc.] embodying these anthropometric extremes. The digital personas were then imported into the system CAD model to evaluate fit and clearance to the lead concepts and system architecture.

Formalised Formative Testing

A formative study was planned involved 25 representative patients over 50 years old, representing a 'worst-case scenario' for usability. This demographic was prioritised based on early clinician insights, which indicated that elderly patients pose the greatest challenges for positioning and fixation during OCT scanning. As ophthalmic care needs increase with age, this group was identified as the largest sub-population that would routinely interact with the system.

Participants were recruited through a third-party agency specialising in medical Human Factors research. The study specifically included individuals with:

- Aged 50+, distributed across four age brackets (50–59, 60–69, 70–79, 80+ years).
- Self-reported or medically diagnosed upper back, neck, lower back, or mobility impairments.
- A balanced male-female ratio (~50:50) to reflect general patient demographics.
- A range of mobility restrictions, including two participants who required occasional or frequent wheelchair use.

Since the study's objective was to explore interface feasibility rather than evaluate clinical performance, vision status and prior OCT experience were not required for participation.

Ethnicity was not a formal recruitment factor, as incorporating additional demographic quotas would have significantly increased cost, complexity, and time constraints. While facial morphology is known to vary across ethnic groups, this study aimed to provide early validation of design feasibility rather than a fully representative solution. Although a digital anthropometric analysis (via PeopleSize and CharacterCreator3D) was conducted to refine interface dimensions, it did not influence participant selection. In later phases, broader demographic testing would be recommended to refine the interface further.

Within the elderly patient population, the most problematic cases for opticians and ophthalmologists typically involve:

- Limited neck and upper body flexibility, making it difficult to maintain the required scanning position.
- Difficulty achieving stable eye fixation, exacerbated by postural discomfort.
- Mobility impairments, increasing setup complexity and adjustment time.

To address these challenges, participants were selected to span a range of physical capabilities, ensuring the interface was tested under realistic, demanding conditions. While the primary focus was on elderly users, limited testing with paediatric participants was also conducted in earlier fail-fast design cycles.

To facilitate representative testing, the team developed a functional rig with interchangeable interfaces and adjustable physical parameters, such as height and approach angle. The rig also incorporated a 4K camera (positioned off-axis via a prism) and an inline fixation target to measure patient stability. A custom machine vision application was developed to analyse patient stability during testing, addressing the client's requirement to evaluate the likely impact of interface designs on image quality and likelihood of needing re-scans.

The study evaluated three designs: Silicone Unibody Goggles, 3-Point Rigid Goggles, and the Conventional Head & Chin Rest as baseline to benchmark against. A randomised trial design captured both objective metrics (e.g., anthropometry, eye openness & movement) and subjective feedback (e.g., comfort ratings, interface preference).

Results

Rationale for Metrics & Evaluation Methods

This study aimed to assess whether novel patient interfaces could reduce patient 'chin time', thereby improving throughput without compromising usability or clinical performance. Three key evaluation areas were identified:

- **Setup Time** – The time required for a patient to self-seat in a comfortable and stable position. Since patient positioning contributes significantly to total chin time, this metric was critical in assessing efficiency. Patients were allowed to independently adjust their chair and table height to optimise their position. Setup time was measured as a single span from first interaction with the system to a self-reported 'ready' state.
- **Stability & Comfort in Use** – A user's ability to remain stable during scanning, while also maintaining subjective comfort. This was assessed using a 7-point labelled Likert scale, supplemented by qualitative participant feedback to capture individual experiences.
- **Eye and Head Fixation Stability** – Ensuring that novel interfaces did not compromise fixation quality, a key factor in scan integrity. Stability was quantified by tracking eye motion and head movement over a 30-second window while participants focused on a stationary fixation target. Machine vision methodologies were implemented in MATLAB, analysing point movement to determine variability.

Each metric was recorded for two novel interface designs and a benchmark chin-and-headrest system, tested across multiple inclination angles.

These methods provided structured, quantifiable evidence supporting a 'reason-to-believe' that novel designs could improve efficiency while maintaining usability and scan integrity.

1. Stability and Comfort

The Unibody Goggles were overwhelmingly preferred for comfort due to their softer, more supportive design. However, some participants noted tightness across the nose. Comparatively, the baseline design allows participants to rest into the interface, relieving neck pressure when positioned correctly, but setup was more complex and time-consuming.

Stability assessments using a custom MATLAB machine vision program revealed the Unibody Goggles achieved comparable eye fixation and lower overall head motion compared to the baseline. The 3-Point Goggles showed significantly greater motion and variability, indicating poorer stability. Furthermore, these performed poorly in maintaining 'eye openness', likely due to geometric issues.

Fixation analysis revealed the Unibody Goggles matched the Baseline in mean scores and standard deviations for net eye motion, demonstrating equivalence in visual fixation.

2. Effect of Tilt/Incline Angle

Participants preferred low tilt angles for Goggles interfaces (10–20°) however, higher angles (beyond 30° for Goggles and 10° for the Baseline) were deemed unacceptable due to discomfort, correlating strongly with reported back/neck pain. There were no significant differences in preferences by age, gender, or physical impairments.

3. Set-up time

Statistical analysis indicated no significant difference in setup times between the Unibody Goggles and 3-Point Goggles, but both outperformed the Baseline, which required more frequent camera focus adjustments, suggesting less positional repeatability. Self-adjustable interface height was universally appreciated, enabling participants to fine-tune their position for optimal comfort.

Key takeaways

1. **A Unique, Hypothesis-Driven Approach**: By prioritising a commercial imperative over pure ergonomics, the study demonstrated how novel patient interfaces can improve both efficiency and comfort in OCT systems.
2. **Design Thinking and Simulation Accelerated Development**: A design-thinking methodology, supported by human factors and ergonomic principles, leveraged simulation tools to validate concepts quickly and cost-effectively.
3. **Multidisciplinary, Data-Driven Collaboration**: Success hinged on a multi-disciplinary approach integrating expertise across optics, engineering, software, and human factors to develop and prove out a concept that balanced clinical and business needs.

Usability of Drug Delivery Devices: Current Challenges and Innovative Methods

Cornelia Kratzer & David Grosse-Wentrup

Design Science

SUMMARY

Healthcare human factors are ever evolving, driven by patient safety, technological advancement, and regulatory requirements. While this progress challenges current practices, it simultaneously serves as a driver for innovation. These current challenges are presented, alongside the innovative methods of force studies, injection and hold time measurements, and iterative instructions for use (IFU) design studies, which can be used to address these challenges.

KEYWORDS

Combination products, force studies, injection time measurement, hold time measurement, iterative IFU design

Introduction

In recent years, the landscape of human factors (HF) for drug delivery device development has been shaped by advancements in patient safety, evolving technological innovations, and increasingly stringent regulatory requirements. This development is exemplified by the U.S. Food and Drug Administration's stricter designation of critical tasks, as well as by the demand to record injection time measurements during validation studies. This regulatory rigor, while essential for safeguarding public health, introduces significant challenges not only during device development, but especially when validating device usability.

This paper focuses on three methods to address these challenges: injection force measurements, hold and injection time measurements, and iterative IFU studies. These methods enable precise, iterative, and comprehensive usability assessments, enhancing testing accuracy, efficiency, and effectiveness, ultimately leading to more reliable device performance and better patient outcomes.

Currently, formative usability testing is typically performed linearly, where the results of a study are analysed only after testing is completed, leading to longer durations between studies and fewer iterations and tests of design elements. Injection time measurement, which is used to validate complete dose administration, is often performed with a stopwatch or by analysing video recordings, causing inaccurate measurements as well as requiring extensive time to generate data through video analysis. Lastly, injection forces and other related factors such as injection angle and movement of the injection device or tremor are not typically assessed at all, despite their potential impact on usability and patient safety.

This paper presents these three methods that address challenges related to HF testing of the user interface of drug delivery devices.

Methods

This section outlines methods for evaluating user-applied injection forces, measuring injection and hold times to collect precise administration data, and using iterative approaches to optimise IFU design. Specialised systems were developed to support these methods, facilitating data collection and analysis. The methods were adapted and refined to meet Design Science's specific requirements of usability studies, evolving through repeated application. Their use in proprietary studies provided supporting data on their effectiveness.

Force Studies

The growing number of biologic medications, often characterised by larger volumes, higher viscosities, and longer injection times, places greater physical demands on users of hand-held injection devices. This applies for example to the force necessary for depressing syringe plungers or holding autoinjectors in place on the injection site, as well as maintaining the respective device at a precise angle with little movement or tremor throughout the injection process.

To assess patient capabilities and determine design requirements for the device during formative studies, injection devices and injection pads were equipped with adapters containing load cells and accelerometers, respectively. These modifications were made with minimal impact on the device user interface, ensuring that the modified devices closely resembled the commercial versions. The underlying technology can be easily adapted to other devices, with only minor modifications, offering a versatile solution for a wide range of injection systems and other devices.

During usability testing, the modified devices and injection pads were used in the same manner as the original device, with the injection pad attached to the body. The combination of sensors recorded force data, e.g., plunger depression force or autoinjector depression force, as well as absolute device orientation data, i.e., device orientation and movement, at a sampling rate of 50 samples per second with a tolerance of 1%, ensuring accurate measurements. This level of accuracy ensures that even subtle variations in device use, such as slight changes in force or orientation, are captured with precision.

Injection and Hold Time Measurement

To accurately assess the completeness of injections in autoinjectors, an automated, sensor-based timing system was employed. This system utilised force sensors within specialised injection pads and auditory feedback mechanisms to precisely capture injection and hold times. The auditory feedback system used microphones to detect the click sounds produced by the device at the start and completion of the injection. Audio filtering was applied to block background noise, isolating the clicks generated by the autoinjector during the injection process for further timing analysis.

Additionally, a needle sensor integrated into the injection pad detected when the needle entered and exited the injection pad. These measurements were synchronised with the auditory feedback as well as force data from sensors in the injection pad, providing precise, real-time data for both injection time and autoinjector hold times. While the system used a specialised injection pad, the autoinjector itself remained unmodified, thus preserving its integrity. This makes the system particularly valuable for validation studies requiring commercial-equivalent devices. Overall, this approach provides significant advantages over manual injection timing methods across various research stages.

Iterative IFU Design

To expedite and refine the design of IFUs, several iterative methods were employed during usability studies. First, experts conducted heuristic evaluations prior to usability testing to identify potential issues based on established design principles, allowing for early-stage refinements. Next, designers observed participants from an observation room during study execution. With only small sample sizes most use errors were detected, allowing designers to identify trends and translate observations into IFU design updates in real-time. These updates were then assessed during subsequent sessions of the same study, making the process both efficient and time effective.

By incorporating these strategies within a single formative study, iterative updates, and evaluations of the IFU could be completed within days or weeks, reducing time, and ensuring that design improvements directly addressed user needs and challenges.

Conclusion

Addressing the new developments and challenges in human factors testing requires innovative approaches to enhance both accuracy and efficiency of usability studies. The three methods presented - force studies, injection and hold time measurements, and iterative IFU design - offer benefits over traditional practices. They not only improve overall user experience but also ensure that device user interfaces are intuitive, safe, and effective for the intended users, while assuring regulatory compliance.

Use of rapid genetic testing equipment – a human factors case study

Rachel Corry

Manchester Centre for Genomic Medicine

SUMMARY

A case study of the successful implementation of a rapid bedside genetic test in neonatal units, to reduce the likelihood of hearing loss in premature and sick babies.

KEYWORDS

Genetic test, hearing loss, implementation, healthcare

Introduction

Approximately 90,000 babies are admitted to neonatal units each year, the most common causes are premature birth or sickness. Many of these babies require antibiotic treatment on admission or soon after to treat and/or protect them against infections. The most common antibiotic used in neonatal units is gentamicin. However, around 1/500 babies have a small difference in their genetic make-up – a genetic variance (MT-RNR1 m.1555A>G) - which means that they can develop severe hearing loss or total deafness after only a single dose of gentamicin. Although there are established laboratory-based genetic tests which could be used to detect this genetic variance, these can take an extended amount of time, which is often too long when rapid treatment is required.

Manchester University and a technology company called Genedrive together created a medical device that performed a rapid point of care test (POCT) for this genetic variant that could provide a positive or negative result within 26 minutes. This rapid response meant that babies could be given the most appropriate antibiotic within the first hour of admission to a neonatal unit.

However, having the technology is only the beginning. The key is to make sure it can be used accurately and reliably, and to ensure that it can be integrated into the established, busy admission process for neonatal units. This was trialled in the NIHR-funded research study Pharmacogenetics to Avoid Loss of Hearing (PALOH) (McDermott 2020, 2021, 2022). Human factors good practice principles were reflected in this process. These are not restricted to good ergonomic device design, but also good project planning, organisation and engagement with key groups and specialists. These principles are described as a case study here.

Effective collaboration

The initial project team were awarded funding to bring together a broad team to trial the implementation of the genetic testing device over two busy neonatal intensive care units (NICUs) in the north-west of England. The team was made up of a wide range of disciplines including subject-matter experts, technologists, end-users and Public and Patient Involvement and Engagement (PPIE) contributors, one of whom was a Human Factors specialist. All members attended the monthly project meetings where project decisions were made, progress was summarised and crucially each member of the project team had an equal voice as the project developed.

In addition to this, additional project participants were involved as the project progressed, including experts in health economics to assess the impact that this device and test could prove to be cost-effective when compared against life-long costs associated with hearing loss.

Clarification of the project aims

It is important that the public can buy into the need for new medical devices and equipment. The diverse team and frequent engagement with patient representatives from an early stage helped to maintain the focus of the study. Engagement with the public also helped to raise the profile of the study, making people aware of the genetic variant and how this study had the potential to help reduce its impact by detecting it early.

Identification of the challenges

Early in the project it was acknowledged that the logistics of fitting a genetic test into the already busy first hour of a baby's admission to the neonatal unit (often referred to as the 'golden hour') was key. Exploring the potential to do this involved a task analysis-style approach to reviewing the admission process, the minimum number of people involved and their responsibilities, what was done when, in what order, and where administering the genetic test could be integrated without impeding this finely tuned process.

There are regularly a team of people involved in a baby's admission to the neonatal unit, carrying out a number of different tasks, from the immediate stabilisation of a baby's breathing, attaching the relevant monitoring devices, taking of blood gas measurements and swabs for MRSA. It was important to demonstrate that this new test could be carried out as part of this suite of tests, without compromising any of the existing, essential activities.

In order to not disrupt the usual admission routine, a task review was carried out to identify the best time for the test swab. This concluded that combining this with the routine skin swabs, such as those for covid and MRSA, felt intuitive as it was a single additional task added to an established and well-practiced process. The timing of these swabs, carried out early in the admission process, allowed time for the test result to be processed so that suitable antibiotics could be administered within the first hour of admission.

Identification of end-users and design of the medical testing device

The genetic test was carried out by performing a cheek swab on the baby, then transferring this sample to a handheld medical device for analysis. The end users of the device were identified (in this case 300 neonatal nurses were trained to perform the test) and a review carried out of the current skills and experiences of the neonatal nurses and how these could be drawn on to make the device as familiar and usable as possible to reduce the potential for error.

A task analysis approach was taken when reviewing the number of task steps needed to carry out the test, how this could be simplified and the process made as intuitive as possible. The design and usability of the genetic testing device was reviewed and revised in response to end-user and Human Factors specialist advice. Some of the initial feedback to the test equipment and process included:

- The need to simplify the information inputted into the testing device before the test, making the information input and layout more intuitive and allowing less opportunity for error
- Consideration of how to manage multiple tests being carried out concurrently
- Potential improvements to navigation within the testing device system
- Providing meaningful labels, from the name of the test equipment, to clear and unambiguous indication of the test results

This review also included the clarity of instructions and operator aids to guide the users as they became more familiar with the system and the training provided.

Teamwork was found to be important (Brown, 2024) in how the tasks were organised, e.g. one nurse would take the swabs and be ‘hands-on’ in carrying out admission tasks, the swabs would be handed to a second nurse who would run the tests and wait for results.

Engagement with the trial and understanding of the importance of the test was cited as a key issue in the success of the trial (Brown, 2024). Neonatal nurses reported that it was important to them that they were not harming the babies by giving them an antibiotic that would damage their hearing. Conversely this also meant that failure of a test could increase anxiety and frustration, increasing the importance of resolving any reported user frustrations during the testing. The majority of nurses interviewed at the end of the trial period reported that they felt that the test fitted well into the admissions process without any delays to administering antibiotics.

Errors encountered and how these were addressed

The study found that it had a failure rate of approximately 20.4%, 17.1% of which was attributed to equipment failure. The reasons for these failures were investigated during the trial period, along with initial operational feedback, and a number of improvements were made, which reduced the failure rate to 5.7%:

- A barcode scanner replaced manual input of patient data, which immediately associated a test with a specific baby and reduced potential for input errors at start of test.
- Five false positives were identified during testing. These babies were treated with an alternative antibiotic and underwent traditional genetic testing to investigate the accuracy of the test. The false positives were found to be a result of incomplete testing cartridge insertion – this was addressed by improving the design of the cartridge to ensure that it would be assembled and inserted correctly before the test could be performed.
- Test fails were found to be predominantly associated with low-signal intensity, which was resolved via modifications to the assay buffer following the initial testing period.
- There were instances of equipment not working as expected and the printer not working due to poor connectivity. The second generation of the testing device improved connectivity, improved the interface and reduced input errors.

Accessible information for patients (in this case, parents of patients)

The testing was carried out by medical staff, but for the purposes of the trial it was important that parents of neonatal babies had accessible information, delivered in a variety of different formats to help them understand what the testing involved. Focus groups were carried out with neonatal parents to develop appropriate information materials and provide the patient-parent perspective on other elements of the process, including the consent process (Dawes, 2024).

After care

Within the study, three babies were correctly identified as testing positive for the genetic variant. In addition to the requirement to an alternative antibiotic, it was recognised that the families with babies who tested positive would need support and advice. The project team included Health Psychologists and genetic counsellors who were able to develop appropriate processes and onward signposting to further support.

Reflection & identification of potential improvements

The project assessed 751 babies, three of whom tested positive for the genetic variant and were given an alternative antibiotic as a result. The test was fully integrated into the admissions process and did not delay antibiotics administration. Throughout the project equipment failure rates were improved through ongoing assessment and fine-tuning to reduce the potential for user-error.

Recognition

The results of the testing have been published and the genetic testing device was conditionally recommended by NICE for use in February 2023 (NICE, 2023). The impact of this testing on neonatal has been acknowledged through a number of awards, including the New Statesman Positive Impact in Healthcare award (2022), Manchester University Making a Difference award (2024) and the Times Higher Education STEM award (2024). Funding has been awarding for a follow-on project rolling out the testing across the UK.

References

Brown G, Warrington N, Ulph F, et al. Exploring NICU nurses' views of a novel genetic point- of-care test identifying neonates at risk of antibiotic-induced ototoxicity: A qualitative study. J Adv Nurs 2024;

Dawes P, Arru P, Corry R, et al. Patient and public involvement in hearing research: opportunities, impact and reflections with case studies from the Manchester Centre for Audiology and Deafness. Int J Audiol 2024;63(2):146–54.

McDermott JH. Genetic testing in the acute setting: a round table discussion. *J Med Ethics* 2020;46:531–2.

McDermott JH, Mahood R, Stoddard D, Mahaveer A, Turner MA, Corry R, Garlick J, Miele G, Ainsworth S, Kemp L, Bruce I, Body R, Ulph F, Macleod R, Harvey K, Booth N, Roberts P, Wilson P, Newman WG. Pharmacogenetics to Avoid Loss of Hearing (PALOH) trial: a protocol for a prospective observational implementation trial. BMJ Open. 2021 Jun 16;11(6)

McDermott JH, Mahaveer A, James RA, et al. Rapid Point-of-Care Genotyping to Avoid Aminoglycoside-Induced Ototoxicity in Neonatal Intensive Care. JAMA Pediatrics 2022;176(5):486–92.

National Institute for Health and Care Excellence (NICE). Genedrive MT-RNR1 1D Kit for detecting a genetic variant to guide antibiotic use and prevent hearing loss in babies: early value assessment. Health technology evaluation HTE6. 10th August 2023.

Utilising principles of visual hierarchy to reduce errors of accidental drug administration

Amol Lotlikar

University College London Hospital NHS Foundation Trust

SUMMARY

Programming errors in infusion pumps pose significant safety risks. Current mitigation strategies rely on human-centred methods such as checklists or two-person checks. Redesigning pump interfaces using visual hierarchy principles could better prevent errors by emphasising critical information during drug selection and confirmation.

KEYWORDS

Medical device, Visual hierarchy, Patient safety

Introduction

Total Intravenous Anaesthesia (TIVA) refers to the administration of anaesthetic agents entirely through intravenous methods, facilitated by infusion pumps with pre-programmed drug delivery models. TIVA is gaining prominence in anaesthetic practice due to its clinical advantages (Johnson, 2017) and its ability to reduce the environmental impact of anaesthesia by lowering the carbon footprint Bernat, 2024).

Previous research (Kirkendall, 2020) has highlighted the potential for error when utilising smart pumps for drug delivery, which can range from wider system and performance errors to issues with the pump interface and design.

In our department we use BD Alaris™ pumps for all TIVA delivered. While TIVA offers numerous benefits, it also presents risks, particularly regarding error during pump programming. A departmental meeting recently highlighted a near miss incident involving the administration of an incorrect drug due to programming errors. This report highlights potential issues with the pump interface that could be contributing to increased error rates and suggests areas for future improvement, specifically regarding the visual hierarchy and design of the infusion pump screens.

Method

To investigate this further, a snapshot survey was conducted among 187 department members, with 47 responses received. The survey findings are as follows:

- 38% had mistakenly inputted the wrong drug, concentration, or model in the last year.
- 11% of these errors led to the incorrect drug/model being delivered to a patient.
- 46.8% expressed concern about the potential for such errors in their practice.

These findings indicate that the near miss was not an isolated event, emphasising the need for further investigation and mitigation strategies. Whilst the response rate was low for the size of department, even a small number of errors can lead to significant harm (Kataoka, 2008) or death, warranting further research into potential improvements in pump design.

Current Mitigation Strategies and Interface Design

At present, the primary strategies for mitigating programming errors rely on human-centred methods, such as checklists or two-person verification of pump programming (Woodward, 2024). However, according to Human Factors principles such as the hierarchy of controls, design-level changes are more effective than human-centred solutions (Kelly, 2023).

Potential Issues:

- **Contrast and Readability:** The current pump screen displays white text on a black background. Research on readability and error reduction (Budiu, 2020) suggests that light text on a dark background, negative contrast polarity, may increase error rates compared to positive contrast polarity. Negative contrast polarity may have been chosen primarily for battery life conservation rather than usability.
- **Drug Programming Interface:** the current interface for the BD Alaris^{TM} pump model follows a consistent design (Figure 1):
  - Once selected, drug and model details are displayed on the right-hand side of the screen and remain there even with the final confirmation
  - Patient demographic data is inputted sequentially on the left-hand side of the screen. Attention is therefore further focussed on the left side to ensure correct demographics are inputted.

Although this design is efficient, it lacks emphasis on the critical confirmation step before drug delivery.

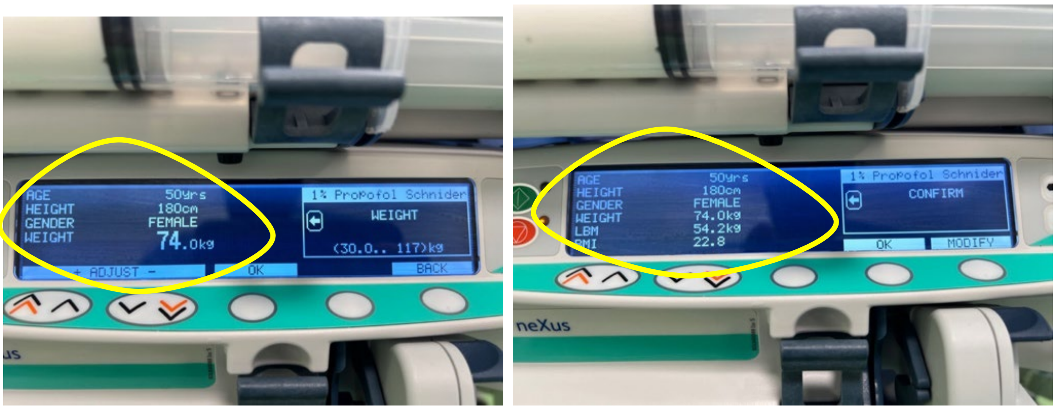


Figure 1: Current drug selection interface.

Proposed interface improvement using principles of visual hierarchy

Visual hierarchy refers to the design principle of arranging information in such a way that guides the user's eyes to the most critical data first. This can be achieved through elements such as size, colour, contrast, and placement (Saw, 2024). With regards to medical pumps, prioritising important details such as drug selection, demographic information and alarm indicators, could ensure quick and accurate decision-making during patient care. Conversely, a lack of visual hierarchy could result in difficulty distinguishing the relative importance of critical information, leading to errors with pump programming.

There is limited research regarding visual hierarchy in medical pumps. However, insights can be drawn from web page user experience studies, where scanning behaviours have been well-documented (Pernice, 2019).

When browsing the web, users tend to scan in patterns to improve efficiency under time constraints and are also influenced by familiarity from previous experiences. It is possible that this scanning method may also occur when programming TIVA pumps due to distractions, alarms, and time pressures. In western languages, users preferentially read text on the left and towards the top of pages as well as focussing on areas that are of interest to them. Consideration of these scanning patterns to pump interface design could provide insights into where errors are occurring. As discussed, with these principles in mind, when programming the BD Alaris™ pump, users focus may be more on the left of the screen which could result in missing if the correct drug and model have been selected.

Proposed steps

To enhance usability and reduce errors in TIVA pump programming, a proposed redesigned interface would include improved contrast and a critical confirmation step before finalising drug delivery. (Figure 2)

This simple measure, of placing the drug name and concentration on the left of the screen, prior to delivering the drug should be trialled to assess its impact on error reduction. Additionally, further work is needed to compare error rates when programming pumps under different contrast settings and in varying lighting conditions.

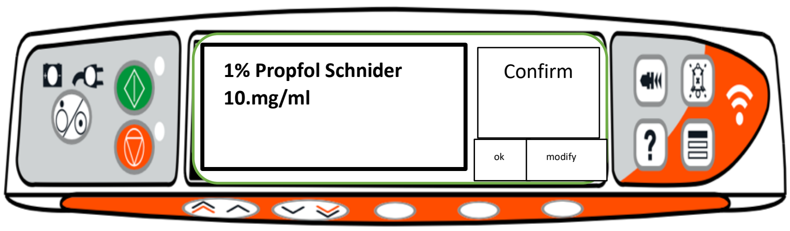


Figure 2: Proposed confirmation step.

This proposal has already been shared with the manufacturer, whose medical device team has acknowledged the suggestion. Findings from future trials should be also shared with patient safety groups for further consideration.

Conclusion

TIVA is a valuable anaesthetic technique with clinical and environmental benefits. However, error in pump programming remains a significant safety risk. By implementing design changes informed by principles of visual hierarchy, it may be possible to reduce these errors and improve patient safety. Collaboration with equipment manufacturers to evaluate these proposed changes and dissemination of findings to professional bodies will provide an important step in identifying design elements that could enhance patient safety.

References

Bernat, M., Boyer, A., Roche, M., Richard, C., Bouvet, L., Remacle, A., Antonini, F., Poirier, M., Pastene, B., Hammad, E. and Fond, G., 2024. Reducing the carbon footprint of general anaesthesia: a comparison of total intravenous anaesthesia vs. a mixed anaesthetic strategy in 47,157 adult patients. Anaesthesia, 79(3), pp.309-317.

Budiu, Raluca. Nielsen Norman Group (2020) 'Dark Mode vs. Light Mode: Which Is Better?', February. Available at: https://www.nngroup.com/articles/dark-mode/ (Accessed:04/02/2025).

Johnson, K.B., 2017. Advantages, disadvantages, and risks of TIVA/TCI. Total Intravenous Anesthesia and Target Controlled Infusions: A Comprehensive Global Anthology, pp.621-631.
Kataoka, J., Sasaki, M. and Kanda, K., 2008. Visual behavior differences by clinical experience and alarming sound during infusion pump operation. Japan Journal of Nursing Science, 5(2), pp.123-129.
Kelly, F.E., Frerk, C., Bailey, C.R., Cook, T.M., Ferguson, K., Flin, R., Fong, K., Groom, P., John, C., Lang, A.R. and Meek, T., 2023. Human factors in anaesthesia: a narrative review. Anaesthesia, 78(4), pp.479-490.
Kirkendall, E.S., Timmons, K., Huth, H., Walsh, K. and Melton, K., 2020. Human-based errors involving smart infusion pumps: a catalog of error types and prevention strategies. Drug safety, 43, pp.1073-1087.
Pernice, Kara. Nielsen Norman Group (2019) 'Text Scanning Patterns: Eyetracking Evidence', August. Available at: https://www.nngroup.com/articles/text-scanning-patterns-eyetracking/ (Accessed: 04/02/2025).
Saw, Jessica J, Gatzke, Lisa P, Designing visual hierarchies for the communication of health data, Journal of the American Medical Informatics Association, Volume 31, Issue 11, November 2024, Pages 2722–2729, https://doi.org/10.1093/jamia/ocae175
Woodward, W., Carrannante, J. and Dua, K., 2024. PRESS (Propofol, Remifentanil, Electricity/EEG, Setup and Setting) to Start: Introducing a Total Intravenous Anaesthesia Checklist at a Large Teaching Hospital. Cureus, 16(3).

Allocation of Function: Yes, no, maybe?

Adrian Wheatley[1], George Charalambous[2]

[1]Wheatley Consulting Ltd, [2]Assystem

SUMMARY

This paper summarises the work undertaken to produce an Allocation of Function method in support of a number of client projects in the nuclear sector. This led to the development of an interactive Excel based tool to support the client in Allocation of Function decision making.

KEYWORDS

Allocation of Function, Function Allocation, Method, Strategy, HF Tools, Nuclear

Introduction

The consideration of Allocation of Function is important during the early stages of a project to ensure human capability is accounted for when assigning functions (especially safety functions) to either humans or machines.

The authors were approached by a number of clients in the nuclear sector requesting assistance in the implementation of an Allocation of Function strategy and method early in the lifecycle of their projects.

Although there are a number of Allocation of Function methods and guides that are recognised as good practice, including those made available from organisations such as British Standards (BS), the International Atomic Energy Agency (IAEA) and Electric Power Research Institute (EPRI), it was identified that a degree of adaptation would be required to derive a simplified method that would meet client specific needs.

The authors undertook to devise an Allocation of Function method that could be easily understood and applied by non-Human Factors (HF) professionals with a view to encouraging the engineering teams to consider human capability during the early design development phase.

Method

The work was undertaken in four key phases as outlined in the following sub-sections.

Literature Review

A review of the primary (most referenced) Allocation of Function guidance and methods was undertaken to develop a broad understanding of the context and good practice.

Learning from Experience (LfE)

The authors identified previous work undertaken in the area by both their employing organisation and other organisations in the nuclear sector, including their clients, to inform the development of a method that would meet the specific client's needs.

Understanding Client Need

It was recognised that both clients were at a similar phase in their project lifecycle, and had similar requirements with respect to the development of an Allocation of Function method. Specifically, the key drivers were to develop a method that:

- Could be applied early in the project to support initial design decision making.
- Did not require an extensive understanding of HF or human capability.
- Provided a simple decision-based model that could be applied by non-HF professionals.
- Was scalable or flexible in its complexity to be further developed or of further use as design details emerge.

Allocation of Function Method Development

The initial literature review and LfE identified four key phases or strategic elements to Allocation of Function, these are:

- Identify Functions.
- Characterise Functions.
- Allocate Functions.
- Validation.

The client needs specifically focussed on the method of characterising and allocating functions.

A set of guidance questions and prompts were developed based on the application of understanding of human capability and the role of the operator in the context of the operation or application of the function.

Results

The final method is presented as a semi-automated flow chart within MS Excel (but can be applied in paper form). Guidance is provided at each decision node in the form of yes/no questions and consideration prompts to support the assessor. The logic of the process flow drives the user toward one of the following determinations.

- Implement Fully Automated Solution.
- Implement Fully Automated Solution with Supplementary Monitoring Information Provision.
- Implement Fully Automated Solution with Manual Response to Functional Failure.
- Automation necessary but not feasible. Reconsider functional requirements.
- Manual necessary but not feasible. Reconsider functional requirements.
- Implement Fully Manual Solution.
- Implement Manual Solution with Supplementary Information Provision.
- Implement Blended Solution.

Using the features available within Excel, the tool provides a summary output table in the form of a Dashboard of fundamental factors that influenced the final determination.

Allocation of Function Method

Figure 1 provides an illustration of the decision tree that is used to support options analysis. At each decision box, a number of additional prompts and questions are referred to in support of decision making, these are set out in Tables 1 - 4.

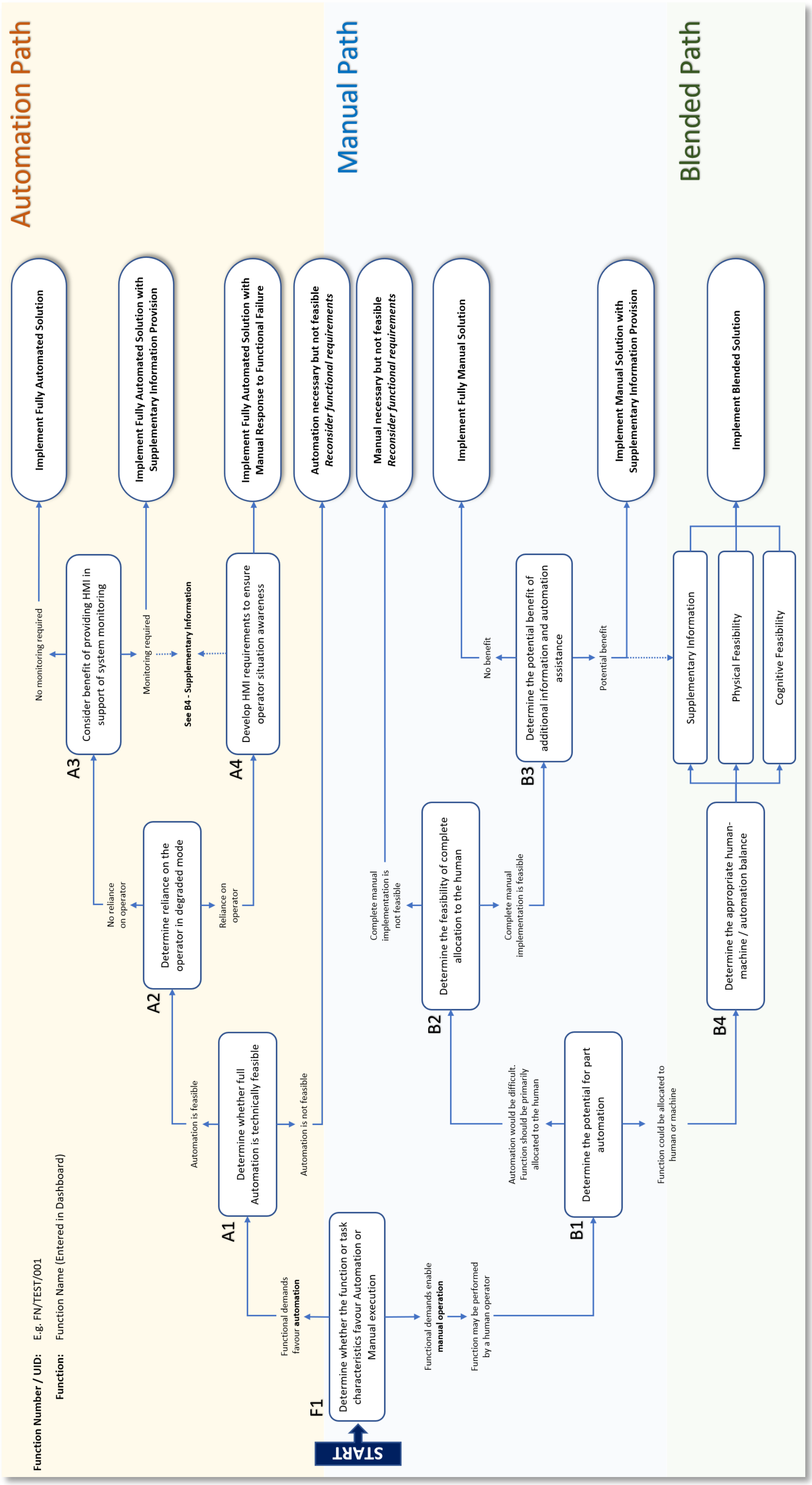


Figure 1: Allocation of Function Decision Tree

Table 1: Primary function prompts

F1: Primary function characteristics or task demands (that would favour automation)	
Legal or statutory requirement	High level of reliability required in support of safety
Operation in an environment hostile to humans	Sustained high intensity operation
Sustained application of high forces	Highly repetitive action
Rapid response to signal or event	Multiple parallel physical operations
Levels of precision beyond normal human capability	Significant data acquisition and processing from multiple sources
High levels of parallel information processing	Retention of a large amount of information
Recording or transferring data to a high level of accuracy	Reliable detection of rare signals or events
Detection of signals beyond human capability	

Table 2: Automation path prompts

A1: Technical feasibility (positive response favours automation)
Are there proven technologies available that can perform this function?
Would the potential cost of developing a bespoke system be acceptable / practicable?
Are anticipated or predictable through-life costs acceptable or sustainable (maintenance | upgrade | replacement)?
Could a solution be implemented within the required timescales?
A2: System robustness (positive response favours automation)
Is there a suitable level of redundancy in the system?
Are (non-human) back-up systems in place?
Are (non-human) further lines of defence in place?
A3: Monitoring requirement (positive response favours supplementary information)
Would monitoring capability favour operator Situation Awareness?
Would monitoring capability favour operator skill-fade?
Would monitoring capability favour operator reaction / response time?

Table 3: Manual path prompts

B1: Technical feasibility (positive response favours manual solution)
Management of novel or unexpected events?
Dynamic adaptation to ever-changing conditions?
Use of heuristics, investigative skills or inductive reasoning?
Decision making based on incomplete or degraded data or information?
Extensive maintenance?
Communication and co-ordination between personnel (on and off site)?
B2: Fully manual feasibility prompt questions (positive response favours manual solution)
Is the function within the physical capability of the operator?
Is the function within the cognitive capability of the operator?
Is the potential human resource cost acceptable?

Table 4: Blended path consideration prompts

B4: Supplementary information consideration prompts

Task prompting.
Task direction / instruction.
Assistance in acquiring information.
Assistance in processing information.
Tasks status information.
Tasks success / failure feedback.
B4: Physical feasibility consideration prompts
Manual handling of large, heavy, complex or fragile items.
Frequency of operation - potential for fatigue.
Poor reach / access / postures.
Duration of effort - potential for fatigue.
Impact / shock / vibration.
Environmental (Noise, light, temperature, humidity, air quality, radiation).
Remote control capability.
B4: Cognitive feasibility consideration prompts
Need to collect data from multiple sources.
High information processing burden.
Frequency of operation - potential for fatigue.
Duration of effort - potential for fatigue.
Parallel information processing.
Potential difficulty perceiving information.
Reliance on memory, retention and recall of information.
Temporal Aspects - absence of time available.
Potential for boredom.
Vulnerable to human biases.
Need for complex error handling.

Application Notes

During the development and early application (trials) of the methodology, the following points were identified that would benefit any practitioner wishing to use, develop or implement a programme of Function Allocation.

- At the start of any large project, especially safety significant projects, a campaign of education and training may be required (typically initiated by HF) to raise the profile and awareness of the need to consider Allocation of Function during optioneering and design development.
- It is important to recognise that although the consideration of Allocation of Function is often championed or led by the HF team, the work requires recognition and collaboration from a number of key stakeholders, including engineering, safety, operations and maintenance. The output cannot be delivered by HF alone and in many cases, it may be preferable for the engineering function to take ownership of the process, recognising the need for support from a HF professional.
- With modern ways of working on large projects, and geographically distributed teams, it is often impractical for professional HF resource to be present at all design reviews or support all optioneering assessments. Awareness training and delegation of responsibility becomes increasingly important, to establish the adoption of a strategy and methodology that encourages consideration of Allocation of Function (and human factors) proportionately. Although there is a risk that HF professionals may have a diminished role in supporting

design decision making, the consideration of human factors (even by others) is fundamentally more important.

- It is tempting to rush ahead with the development and implementation of a method to consider Allocation of Function without first developing and implementing a strategy that determines how the Functions are identified and characterised. This will require collaboration between the key project stakeholders. Many (large) organisations will have their own methods through which Functions are identified (e.g. Functional Analysis, Functional Decomposition, Hazard Analysis, Systems Engineering, Task Analysis). Allocation of Function must be aligned and integrated with the larger programme of work. Specifically, Allocation of Function as both a concept and method needs to be embedded with (or at least used alongside) the core and / or established organisational design decision making processes and arrangements.
- There are many different aspects to Allocation of Function, and it is recognised that there is not one best method that suits all contexts. All models will have limitations and it is likely that any specific approach set out will be adapted for use. The method proposed can be used and / or adapted as required, but provides a robust basis for development, focussing on identifying the key aspects (and human performance shaping factors) to be considered as prompts.
- Proportionality and timing are important aspects to Allocation of Function that are often linked. At the start of the project, specific details may not be available, and the Functions may be loosely defined. However, the consequences of error may be no less severe, and the importance of considering human factors at an early stage may be pivotal to design progression. Once the designs are more developed and specific Safety Functions are being defined, the method of Allocation of Function may need to be adapted, and consideration of human factors should be embedded in the decision-making process.
- In line with design development and option selection, consideration of Allocation of Function may need to be undertaken iteratively, as information becomes available.
- A key aspect to the adoption of an Allocation of Function method is the improved level of rigour and robustness it provides. Through the consistent use of a systematic approach, the rationale for decision making can be captured – which is often poorly documented. The Allocation of Function method should include the means of ensuring that all decisions are captured and recorded.
- There is often a temptation to 'score' or add a quantitative element to the assessment of design options. Although there are valid methods of doing this it often adds unnecessary complexity and ambiguity (e.g. how to validly apply weighting to each parameter). The focus of Allocation of Function assessment should be qualitative.

Summary

The Allocation of Function method developed, and Excel-based tool provides the user with a simplified means of considering human capability in order to make a determination early in the design of the system as to whether the function should primarily be allocated to the human, the machine or a blended approach is required. The methodology is adaptable and flexible enough to be applied proportionately, depending on the stage of the project. The tool provides a mechanism for the decision-making process to be clearly recorded.

Attention Capture Experimental Paradigm for Cross-Screen Interaction in Nuclear Power Monitoring System

Wu Xiaoli, Li Yiqun & Li Qian

Nanjing University of Science and Technology, China

SUMMARY

This study analyses the mechanism of operators' cross-screen interactive behaviour during the execution of typical monitoring tasks from the perspective of attention capture, in order to enhance the superiority of attention capture for critical task information on interfaces, ultimately improving operators' efficiency in manipulation. The connection between the operator's perceptual cognitive process and cross-screen interactive behaviour is established through the bottom-up and top-down attention capture mechanisms. A cross-screen interaction experimental paradigm is proposed, and experiments are conducted in specific scenarios to verify and refine the potential behavioural mechanisms. The experimental results will help to gain a deeper understanding of the underlying mechanisms of cross-screen behaviour and cognitive processes and reveal the mapping relationship between interface factors and cross-screen interactions.

KEYWORDS

Attention Capture Mechanism, Nuclear Power Monitoring, Cross-Screen Interaction, Experimental paradigm

Introduction

The advancement of intelligent nuclear power technology has intensified human factors challenges in main control rooms (MCRs), particularly regarding multi-screen interactions (Xue, 2015). As shown in Figure 1, the MCR is the "nerve centre" of the nuclear power display and control system and is also a typical multi-screen interactive scene (Zhang et al., 2019). The most common human-computer interaction behaviours is the cross-screen information search activity that occurs on different screens or paper documents (Wu et al., 2018), as shown in Figure 2. Operators must manage information distributed across visual display units (VDUs), large display panels (LDPs), and paper documents at key information centres (KIC), leading to cognitive overload risks including visual fatigue, attention diversion, and decision-making errors. This paper investigates the relationship between operators' cognitive processes and cross-screen interaction behaviours during critical tasks. We propose visualisation methods that optimise interface design through attention screening mechanisms, suppressing non-essential information while enhancing task-relevant data processing. Our approach addresses three key design considerations: 1) Information configuration tailored to specific operational scenarios 2) Feedback mechanisms supporting cross-screen information searches 3) Interactive operations mitigating the cognitive impacts of intelligent system complexity. The research aims to improve information comprehension and reduce human error in next-generation nuclear power control systems.

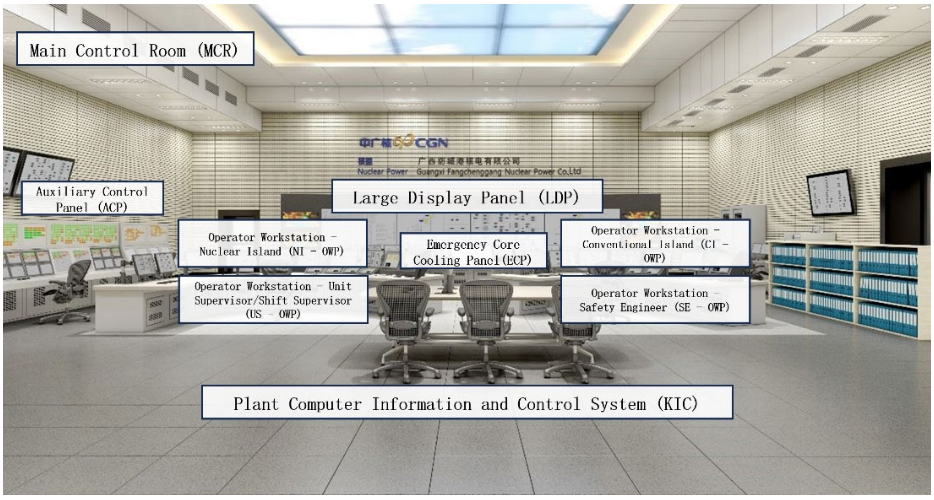


Figure 1. Nuclear power control room layout and working scenes

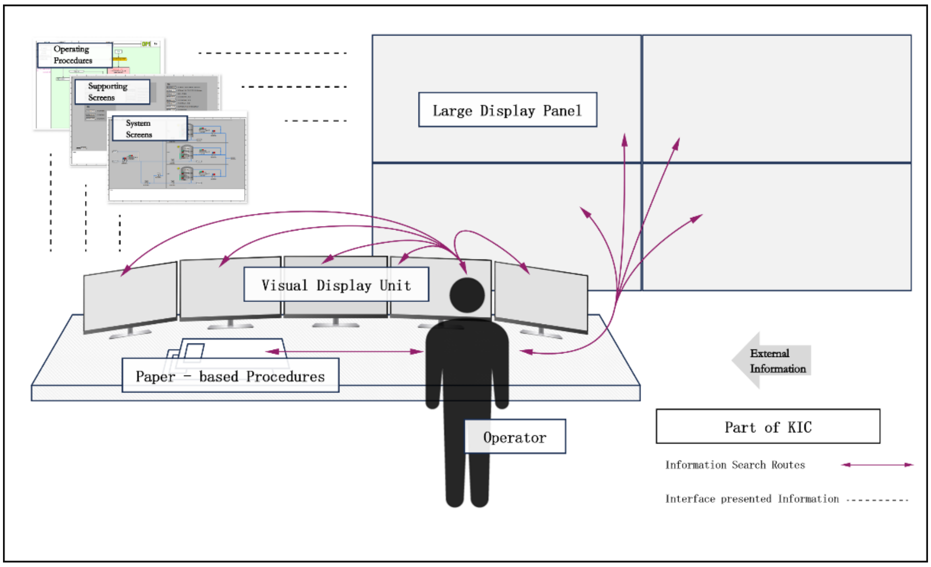


Figure 2. The cross-screen information search activity

Interaction between screens can be divided into same-screen interaction and cross-screen interaction. Cross-screen interaction can be further divided into cross-screen interaction within the same device group and cross-screen interaction across device groups (Nebeling et al., 2013). In the MCR, an operator interacts with multiple VDUs, LDPs and other device groups. Individual cognitive processes guide the attention capture of visual perception. This interaction causes operators to produce individual behaviours based on their own cognitive characteristics. Therefore, the operator's cross-screen interactive behaviours and cognitive process show a close correspondence, involving the interaction between information elements and cross-screen contexts, as shown in Figure 3.

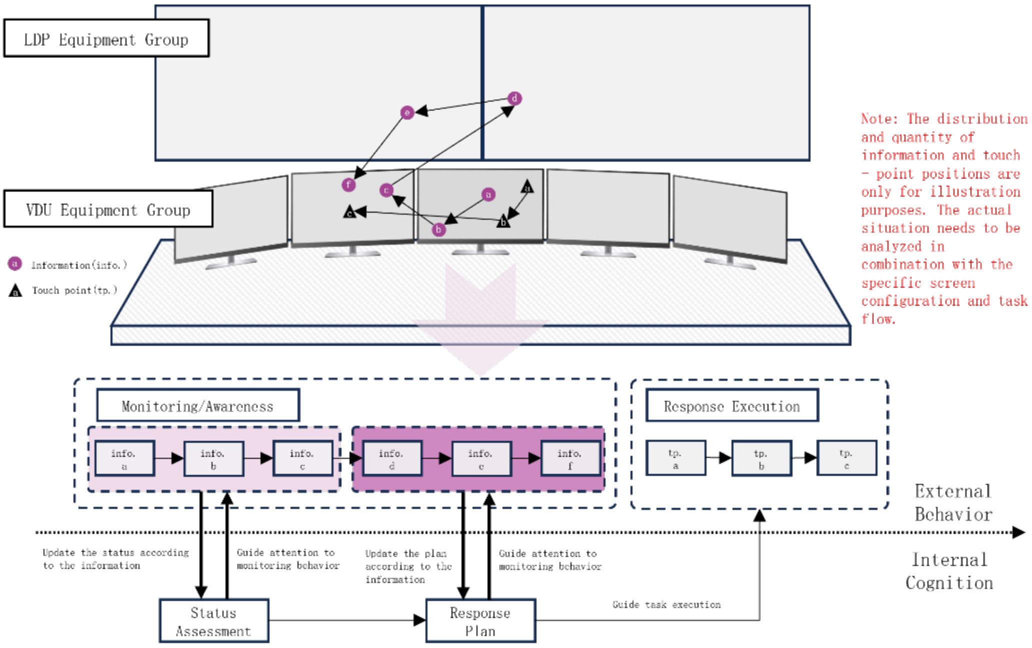


Figure 3. The correspondence between the operator's cross-screen interactive behaviours and cognitive process

The cognitive process of the operator's cross-screen interaction behaviours not only involves the processing and transmission of information, but also deeply reflects the operator's cognitive strategy and behaviours pattern in a specific task environment. In order to deeply understand this behaviours pattern, this process is analysed from the perspective of the attention capture mechanism and a cross-screen interactive behaviours model of the operator based on the attention capture mechanism is constructed, as shown in Figure 4. This model analyses cross-screen interaction behaviours from the perspective of bottom-up and top-down attention capture mechanisms (Posner, 1980), while integrating the interaction between information elements and screen configurations, providing a theoretical basis for subsequent research on cross-screen interaction behaviours mechanisms.

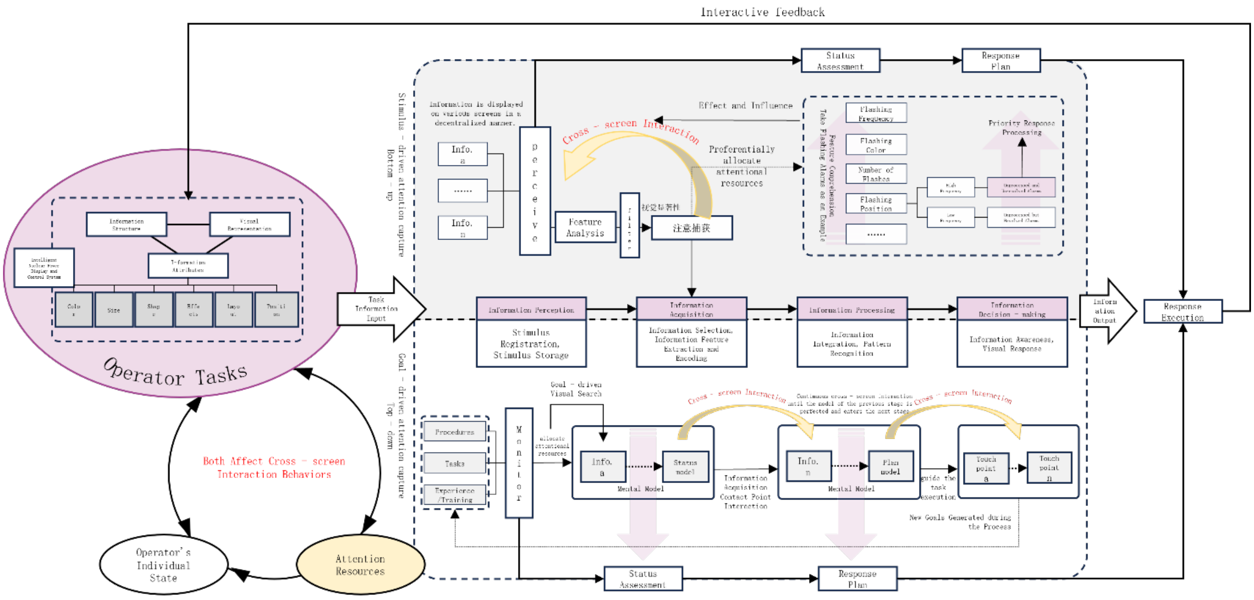


Figure 4. The cross-screen interaction behaviours model based on attention capture mechanism

This chapter details the attention capture mechanism, analysing the operator's cognitive process and cross-screen interaction behaviours. It highlights the importance of this mechanism in cross-screen interactions and proposes a model that integrates the operator's cognition, attention capture, and cross-screen characteristics. The model explains how operators engage in cross-screen contexts influenced by attention capture, providing insight into their behaviours and laying the foundation for experimental paradigms and design.

Methods

Experimental paradigm

Based on the above cross-screen interaction behaviours analysis, a behavioursal experimental paradigm is proposed to study the attention capture effect of operators in cross-screen interaction scenarios. The effectiveness of cross-screen interaction between subjects on different screens is explored from the perspective of attention capture mechanism, as shown in Figure 5.

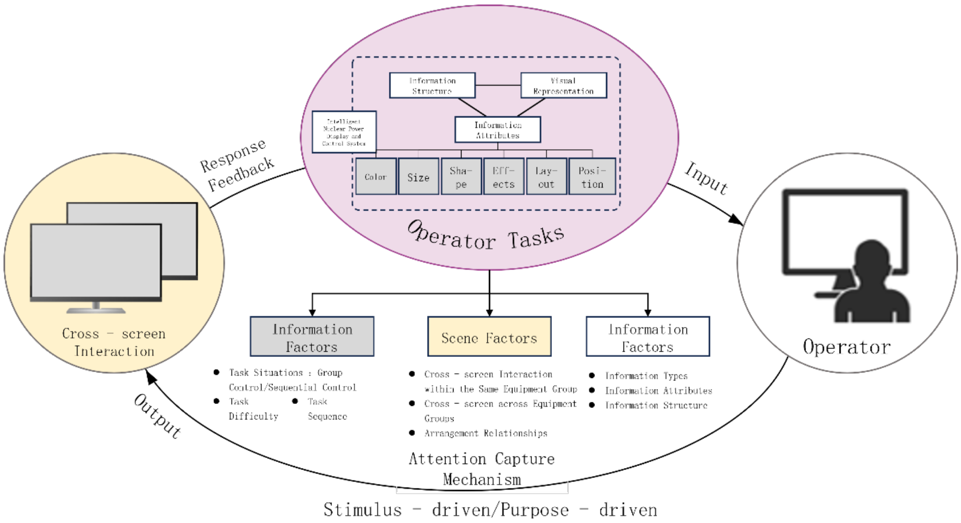


Figure 5. Cross-screen interactive attention capture experimental paradigm under typical tasks of intelligent nuclear power

Experimental design

This experiment uses a cross-screen interaction attention capture paradigm with five VDUs on the operating table. The task involves inserting fault alarm messages during the abnormal diagnosis phase to enhance attention. Psychological, behavioursal, and eye movement data are collected on reaction times, accuracy, and attention-related eye movements (Cheng et al., 2017).

The experimental variables are cross-screen interaction (scene factor) and fault alarm (information factor). Cross-screen interaction includes the number of screens and their orientation, with screen C as the main task starting point. The screen layout and numbering are shown in Figure 6. Fault alarms are defined by three sub-variables: the number of alarms (1-3), alarm level (high: 2 Hz fast flashing, low: 0.5 Hz slow flashing), and alarm location (nine grid positions). The fault alarm form is flashing.

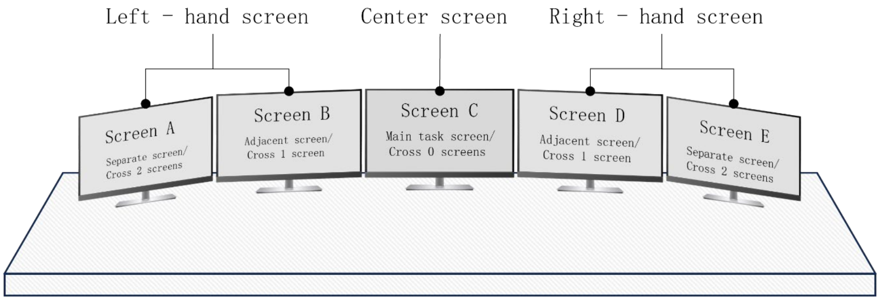


Figure 6. Experimental scene design and screen numbering

Procedures

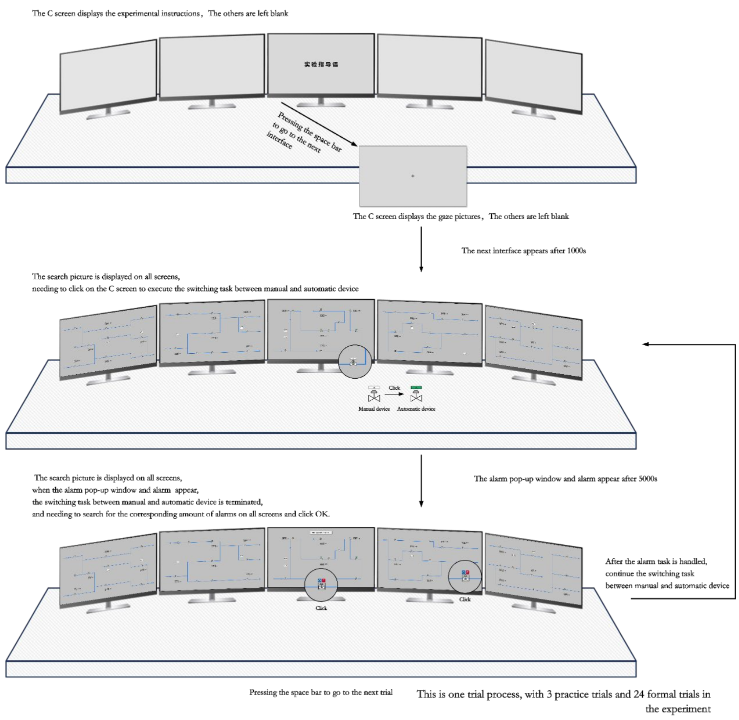


Figure 7. Experimental process diagram

Each subject completes 24 trials with random combinations of independent variables. Before starting, the experimenter provides instructions on a C screen, followed by practice trials. The formal experiment begins with a fixation point ("+") displayed for 1000ms, signaling the start. In the group control task, subjects identify and switch 5 devices from manual to automatic mode by clicking the left mouse button. An alarm pop-up appears during the task, and subjects must pause to locate and confirm the corresponding alarm icons. After handling the alarms, they continue switching the devices until the task is complete. Pressing the space bar starts the next trial. The experimental process is shown in Figure 7. After the experiment, the subjects were required to complete a questionnaire on basic demographic information, including name, gender, age, and years of work experience.

Equipment and Participants

The prototype tool, built in Figma, is tested on an alternating platform using Python's pyautogui and pynput libraries to gather event data. Tobii Glasses, with two eyeglasses, were used, requiring 2-5 minutes for setup before testing. Twelve subjects with experience in nuclear power main control room interface design participated in the experiment, held at the State Key Laboratory of Nuclear Power Safety Monitoring Technology and Equipment, China Nuclear Power Engineering Co., Ltd.

The experimental scene is shown in Figure 8. The subjects, aged 29-44 (average age 35.75), had 4-18 years of work experience (average 7.5 years). All had corrected vision above 1.0 and no color blindness. The experiment used five computers with identical specifications, each displaying 2880x1800px resolution at 64-bit color quality. Subjects sat 55-60cm from the screen.

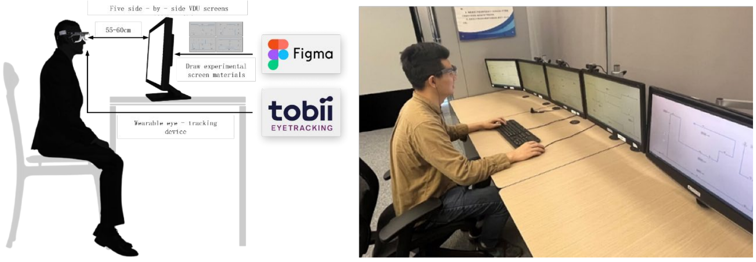


Figure 8. Same device (VDU-VDU) group experiment scenario

Results

Behavioural indicator data

Behavioural data from 12 valid groups underwent Z-score normalisation, with trials exceeding $\pm 3\sigma$ (0.5% of data) removed to eliminate outliers. The data were imported into SPSS for two-way multivariate variance test.

From the Sig values of each significance test, the effect of the cross-screen number on the reaction time is significant ($p = 0.00 \leq 0.05$). The effect of the cross-screen orientation on the reaction time is not significant ($p = 9.00 > 0.05$), and the effect of each factor on the accuracy rate is not significant ($p > 0.05$); The impact of the alarm level on the reaction time is significant ($p=0.00 \leq 0.05$). The impact of the number of alarms and the alarm location on the reaction time is not significant (the number of alarms $p =3.20 > 0.05$; the alarm location $p =0.27 > 0.05$), and the impact of each factor on the accuracy rate is not significant ($p > 0.05$).

By applying the multiple linear regression method, the construction of the multiple linear regression model of the alarm level and the number of cross-screens on the reaction time is statistically significant ($F=76.498$, $P< 0.05$). As shown in Table 1, the regression coefficients of the alarm level and the number of screens crossed are both statistically significant ($P < 0.05$). The coefficients of the number of screens crossed are 2.4851*10-5, indicating that the more screens crossed, the longer the reaction time and the more difficult it is to have an attention capture effect. The alarm level is a binary variable, with a low level coded as 0 and a high level coded as 1. The coefficient is -3.2762*10-5, indicating that a high alarm level has a negative impact on the reaction time and the attention capture effect is more significant. The relationship between the reaction time and the influencing factors of the alarm level and the number of screens crossed can be expressed by the multivariate linear regression model shown in Eq. (1):

Table 1: Multivariate Linear Regression Coefficients for behavioural Indicators

Model		Unstandardised coefficients		Standardised coefficient		
		B	Standard error	Beta	t	Significance
1	(constant)	1.2206E-04	.000	-	21.686	.000
	Alarm level (AL)	-3.2762E-05	.000	-.230	-6.186	.000
	Number of cross-screen (NCS)	2.4851E-05	.000	.403	10.813	.000

Dependent Variable：RT

$$RT = 2.4581 \times 10^{-5} \times NCS - 3.2762 \times 10^{-5} \times AL + 1.2206 \times 10^{-4} \quad (1)$$

Eye movement indicator data

Eye movement data from 12 participants (3 excluded due to low sampling rate) were collected using Tobii Glasses 2, recording number of fixations, duration of fixations, number of saccades, pupil diameter, and scanning path. Data were processed in Tobii Pro Lab with I-VT gaze filtering and aligned to trial-specific time-of-interest (TOI) windows. Outliers beyond $\pm 3\sigma$ (0.5% of data) were removed, consistent with behavioural data protocols. The data were imported into SPSS for two-way multivariate variance test.

From the Sig values of each significance test, the number of cross-screens has a significant effect on the number of eye saccades ($p = 0.05 \leq 0.05$), and has no significant effect on the number of fixations, fixation duration, and pupil diameter data (number of fixations $p = 0.25 > 0.05$; average pupil diameter $p = 2.39 > 0.05$; fixation duration $p = 3.91 > 0.05$). The cross-screen orientation had no significant effect on the eye movement indicators (the number of saccades $p = 1.37 > 0.05$; the number of fixations $p = 1.56 > 0.05$; the average pupil diameter $p = 8.90 > 0.05$; the duration of fixations $p = 2.36 > 0.05$); Due to the limited accuracy of eye movement data, the alarm location was removed from the verification of the fault alarm dimension indicators, and only the number of alarms and the alarm level were retained. The alarm level has a significant effect on the number of saccades and fixations (the number of saccades $p = 0.00 \leqslant 0.05$; the number of fixations $p = 0.00 \leqslant 0.05$), but has no significant effect on the pupil diameter data and fixation duration (the average pupil diameter $p = 8.17 > 0.05$; the fixation duration $p = 5.61 > 0.05$). The number of alarms has a significant effect on the number of fixations and fixation duration (the number of fixations $p = 0.01 \leqslant 0.05$; the fixation duration $p =0.00 \leqslant 0.05$), but has no significant effect on the number of saccades and the average pupil diameter data (the number of saccades $p = 0.42 > 0.05$; the average pupil diameter $p = 4.20 > 0.05$).

By applying the multiple linear regression method, the multiple linear regression model of the alarm level and the number of screen crosses on the number of saccades was constructed, which was statistically significant ($F=13.702$, $P \leqslant 0.05$). As shown in Table 2, the regression coefficients of the alarm level and the number of screen crosses were statistically significant ($P \leqslant 0.05$). The coefficients of the number of screen crosses were 4.772, indicating that the more screen crosses, the more saccades, the more difficult it was for the subjects to concentrate, and they kept switching from one stimulus to another, and it became more difficult to produce an attention capture effect on a fixed target stimulus. The alarm level is a binary variable, with a low level coded as 0 and a high level coded as 1. The coefficient is -5.558, indicating that a high alarm level has a negative impact on the number of saccades. Under this condition, the subjects highly focus their attention on the target stimulus and have a higher visual search efficiency. The relationship between the number of saccades and the influencing factors of the alarm level and the number of screen crosses can be expressed by the multiple linear regression model shown in Eq. (2):

Table 2 :Multivariate Linear Regression Coefficients for Eye-tracking Indicators

Model		Unstandardised coefficients		Standardised coefficient		
		B	Standard error	Beta	t	Significance
1	(constant)	9.693	2.027	-	4.783	.000
	Alarm level (AL)	-5.558	1.781	-.244	-3.120	.002
	Number of cross-screen (NCS)	4.772	1.184	.316	4.030	.000

Dependent Variable: Number of eye saccades (NES)

$$NES = 4.772 \times NCS - 5.558 \times AL + 9.693 \quad (2)$$

Discussion

This study used a fault alarm awareness task to examine the effects of cross-screen interaction and fault alarm variables on behavioural and eye movement indicators, including reaction time, accuracy, fixation count, fixation duration, saccade count, pupil diameter, and scanning path. Key findings include that the number of cross-screens, alarm level, and number of alarms significantly influenced these indicators, while cross-screen orientation and alarm position did not.

The number of cross-screens negatively impacted attention capture, while alarm level positively affected it. Multivariate regression models (Eq. (1) and (2)) describe their interaction. Figure 9 shows a positive correlation between reaction time and saccades in cross-screen environments. Reaction times were shorter and fewer saccades occurred with higher alarm levels. As displays exceeded the subject's field of vision, they had to move physically to locate targets, which reduced attention capture and task efficiency, aligning with the findings on cross-screen interaction with multiple devices (Majrashi et al., 2016). High-frequency flashing alarms effectively conveyed urgency, attracting attention and prompting faster, more focused responses.

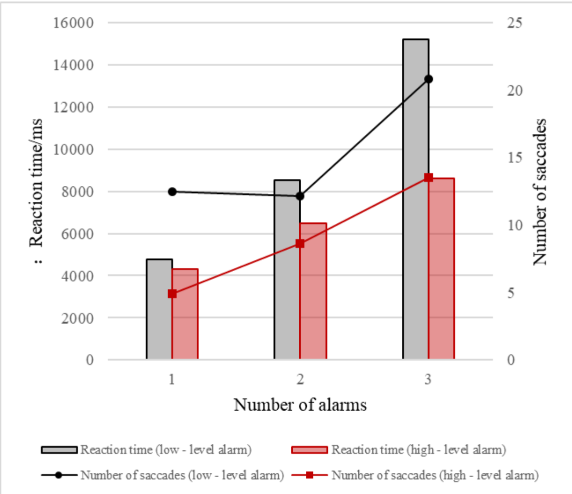


Figure 9. Response Times and Number of Saccades at High and Low Alarm Levels across Different Numbers of Screens

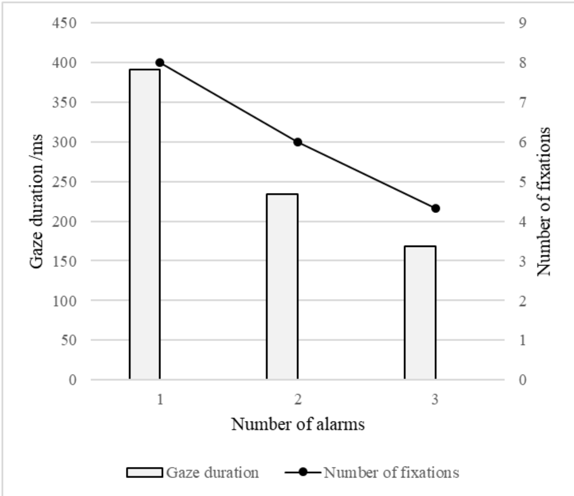
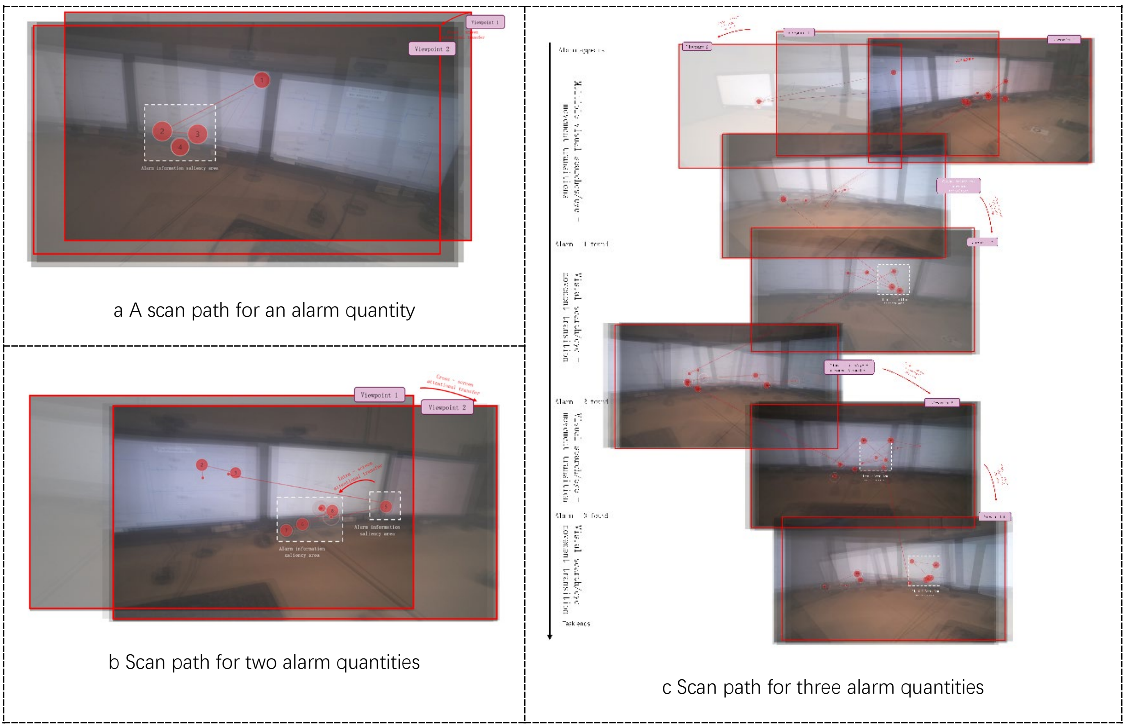


Figure 10. Number of Fixations and Duration of Fixations at Various Alarm Quantities

The effect of the number of alarms on the attention capture effect is mainly reflected in the eye movement indicators related to fixation. As shown in Figure 10, the number of fixations and the duration of fixations both decrease with the increase in the number of alarms, indicating that the more stimuli there are, the more limited the attention capture effect will be. Multiple alarms cause the subjects to be distracted, quickly shift their gaze between many potential targets, and not process a single alarm in depth. This can also be understood as the perceptual similarity between multiple targets. After discovering the first alarm, the subjects can quickly find the remaining alarms without investing too much attention resources. This is consistent with the experimental results in a single screen and is not affected by the cross-screen context (Gorbunova, 2017). The scanning path can clearly represent the annotation behaviours characteristics of the subjects in obtaining target stimulus information when completing the fault alarm visual search task. Even if all subjects are faced with similar search tasks, the scanning path guided by visual attention allocation is affected by individual differences. This difference suggests that individuals adopt different attention resource allocation strategies, or that there are differences in information processing capabilities and experience levels. Therefore, this study selected a typical subject's scanning path under different numbers of alarms for analysis, as shown in Figure 11.



a A scan path for an alarm quantity

b Scan path for two alarm quantities

c Scan path for three alarm quantities

Figure 11. The scanning path of a typical subject under different alarm numbers at high alarm levels

The relevant eye movement data under each alarm number task are shown in Table 3. It can be seen that when the subjects search and respond to fault alarms, the number of fixations and eye saccades increases with the increase in the number of alarms. The fixation points are mainly distributed in the middle of the visual field. When there are multiple alarm information in the visual field, the alarm information on the screen will have a better attention capture effect, and then produce attention transfer within the screen. When the alarm information is not in the visual field, the subjects will actively turn their perspectives for visual search, which is consistent with the point that the subjects need to shift their perspectives and look at the area outside the initial visual field, resulting in a decrease in the proportion of fixation behaviours on the original area (Lavoie et al., 2018). The more fault alarms there are in the cross-screen interaction task, the more cross-screens there are, the more dispersed the fixation points are, and the more complicated the scanning path will be. When the number of alarms is large and distributed across screens, as the subjects turn their perspectives, multiple invalid information search behaviours will occur, affecting the performance of the alarm information perception task. Optimising the layout of the monitoring screen is an effective way to avoid excessive operator shifting of perspective in the cross-screen behaviours scenario design.

Table 3: Eye-tracking data for a typical subject at high alarm level across various alarm quantities

Number of alarms	Number of fixations/times	Fixation duration /ms	Number of eye saccades /times	Average pupil diameter /mm
1	4	412.25	4	3.32271
2	11	303.2727273	14	3.43180909
3	49	424.5102041	77	3.620124082

Conclusion

This research focuses on key cross-screen interaction behaviours within nuclear power monitoring systems, integrating visual attention capture mechanisms. A cross-screen interaction attention capture experimental paradigm tailored to specific task contexts is proposed. The results obtained from the experiments will facilitate a deeper understanding of the underlying mechanisms of cross-screen behaviours and cognitive processes. Additionally, this study aims to reveal the mapping relationships between interface factors and cross-screen interaction, providing empirical evidence for subsequent interface design and information presentation.

Acknowledgments

This work was supported by State Key Laboratory of Nuclear Power Safety Monitoring Technology and Equipment (K-A2021.419), the National Nature Science Foundation of China (52175469), and Jiangsu Province Nature Science Foundation of China (BK20221490).

References

Cheng, S. W., Shen, X. Q., Lu, Y. H., et al. (2017). Eye-tracking based cross-device distributed attention awareness interface. Journal of Computer-Aided Design and Computer Graphics, 29(9), 1713-1724.

Gorbunova, E. S. (2017). Perceptual similarity in visual search for multiple targets. Acta Psychologica,173, 46-54.

Lavoie, E. B., Valevicius, A. M., Boser, Q. A., Kovic, O., Vette, A. H., Pilarski, P. M., ... & Chapman, C. S. (2018). Using synchronized eye and motion tracking to determine high-precision eye-movement patterns during object-interaction tasks. Journal of vision,18(6), 18-18.

Majrashi, K., Hamilton, M., & Uitdenbogerd, A. L. (2016, July). Cross-platform cross-cultural user experience. In Proceedings of the 30th International BCS Human Computer Interaction Conference. BCS Learning & Development.

Nebeling, M., Zimmerli, C., Husmann, M., et al. (2013). Information concepts for cross-device applications. Proceedings of the Conference, 14-17.

Posner, M. I. (1980). Orienting of attention. Quarterly Journal of Experimental Psychology, 32(1), 3-25.

Wu, X. L., Xue, C. Q., Gedeon, T., et al. (2018). Visual search experiments of information features in digital monitoring task interfaces. Journal of Southeast University (Natural Science Edition), 48(5), 807-814.

Xue, C. Q. (2015). Design methods and applications of digital interfaces for human-computer interaction in complex information systems [M/OL]. Nanjing: Southeast University Press.

Zhang, L., Dai, L. C., Hu, H., et al. (2019). Human Factors Reliability in Digital Nuclear Power Plants [M]. Beijing: National Defense Industry Press.

Day in the Life Of: Applying the Process to the Nuclear Industry

Lisa Kelly[1], Rachel Selfe[2] & Suzy Sharpe[1]

[1]Mott MacDonald, [2]AtkinsRéalis

SUMMARY

Day In The Life Of (DITLO) methodologies have been utilised in the rail industry to elicit user experiences and drive a greater shared understanding of how systems or designs will operate in a variety of scenarios. This paper describes the application of this approach to the nuclear industry, in order to improve the integration of design, engineering and safety disciplines, and to achieve a holistic understanding of the user's environment to enhance operability and human performance.

KEYWORDS

DITLO, Scenarios, Workshops, Nuclear, Multidisciplinary

Introduction

Nuclear construction projects typically follow a standard approach to construction, with stakeholders developing designs within separate teams and then presenting these in reviews for wider discussion. Whilst this approach can result in a safe design that meets project requirements, it does not necessarily result in one that recognises the needs of its users and the tasks they undertake. Issues can often be identified too late to enact real change, requiring procedural and other post-hoc solutions to mitigate risks.

On a recent nuclear construction project, Human Factors (HF) Specialists identified an opportunity to improve the collective understanding of a building's future operation by utilising rail industry knowledge of Day In The Life Of (DITLO) methods. This process provides additional benefits to standard construction processes, as it can be used to inform changes to technical requirements, roles and processes, in addition to reducing risk (Bye 2017) and increasing operability.

The DITLOs were undertaken during Royal Institute of British Architects (RIBA) design stages 2 and 3. The goal of the DITLOs was to identify key user needs, requirements, issues, risks, dependencies, assumptions and inform analysis such as task analysis and process flow diagrams.

DITLO Scenarios

Prior to DITLOs being undertaken, the HF team reviewed scenarios to develop a list of potential scenarios or topic areas that demonstrate the design is operable (see Figure 1). The following list was created:

- Normal scenarios – analysis of a normal day, moving through the building and operational tasks therein. Consideration of all users within the building.
- Abnormal scenarios – conditions which occur less frequently but may have a minimal impact on normal operations, e.g. lights requiring replacement, fire drills.
- Degraded scenarios – less frequent conditions with a moderate impact upon operations within the building, e.g. preventative maintenance.

- Emergency – fire or other emergency scenario, including emergency responder tasks, site and building security incidents.

The type and number of DITLOs applied on a project is likely to be dependent on the number of key scenarios used to demonstrate the building is fit for purpose, with further scenarios potentially being identified during initial DITLO sessions. DITLO scenarios are developed in consultation with other stakeholders such as end-users, Operational Readiness and Capability. Example scenarios are presented in Figure 1, and promote discussions in the DITLO scenario definition phase.

Application of the DITLO methodology

DITLO methods can be applied at macro and micro levels: from following a user from the moment they start their shift to the moment it ends, or for one task they undertake within a day. DITLO is a flexible process and bounded by the purpose of why it is being undertaken. For example, the DITLO may want to improve understanding of:

- The detail of particular tasks undertaken at certain points in the day, to understand where safety may be compromised or human errors occur.
- The interaction between the user activities and the physical environment, to recognise whether the design/environment supports the user in effectively completing their tasks.
- Detailing tasks and mapping those tasks, to understand where there may be gaps in procedures and processes.
- Gaining task detail and human movement, to inform other disciplines such as architecture, HVAC and mechanical handling.

The goal of the session requires defining with other stakeholders in advance, to create a framework to bound the DITLO, ensuring it remains focused and that the key project questions to be addressed during the DITLO are captured.

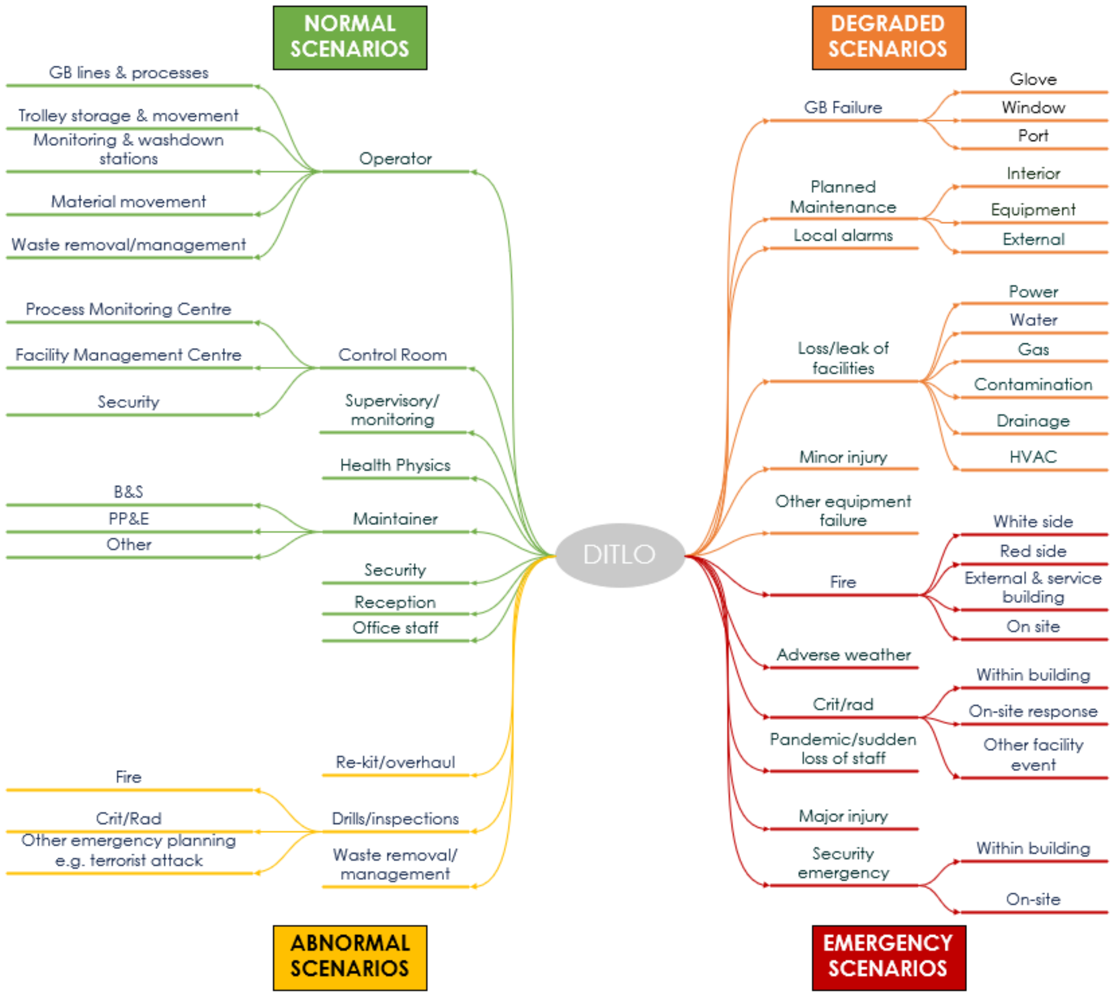


Figure 1: Example DITLO Scenarios

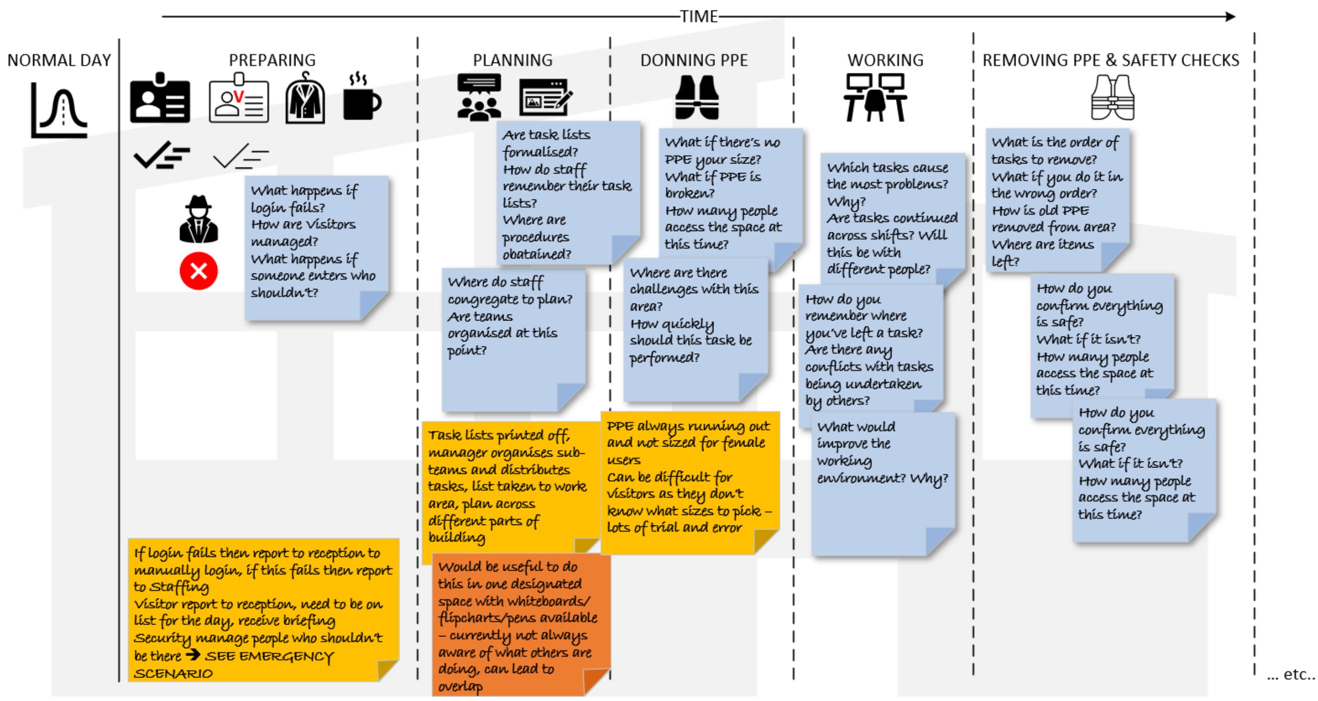


Figure 2: Example swim lane with sticky notes

Method

Preparation

Agendas or briefing notes were provided in advance of DITLO sessions, to ensure the appropriate stakeholders were identified and the scope and boundaries of the session were clearly communicated. Other simple, but necessary, logistical aspects such as the workshop location and room size, including the availability of food and welfare facilities, were carefully planned in advance to provide an inclusive working environment.

Conducting the workshop also required ensuring flipcharts, screens for presenting information and audio equipment were sufficient. This is in addition to planning dynamic data capture methods such as swim lanes, and how they were presented during the day, e.g. the need for sticky notes, marker pens, etc. Planning these aspects ensured the HF team were supported by the environment in delivering the DITLO workshop.

Managing the DITLO

Where possible, the HF team utilised an in-person workshop format to discuss different scenarios, following a multidisciplinary, holistic approach that focused on the detail of day-to-day end user tasks (Bye 2017). This provided flexibility and helped to ensure attendance. The goal was to have all DITLO stakeholders in the room, however, due to the geographic location of project team members and the volume of attendees, a blended approach was utilised, with those contributing the most, such as end-users, being present in the room and other wider project members attending virtually. The scribe, therefore, also monitored the meeting 'chat' function for questions or comments to input into the discussion to ensure key points were not missed. The facilitators ensured that those attending the meeting virtually were included within discussions, providing prompts and opportunities to ensure they had the space to contribute.

During each workshop the following format was followed:

- Define the scope and task(s) being explored in the DITLO – clarify scope of the DITLO with attendees to ensure the discussion remains focused on the topic.
- Present the purpose of the DITLO, clarify role of HF for the day, and allow all those attending to introduce themselves, including name and their role in the project.
- Conduct detailed walk/talk-through of the DITLO topic:
    - Focusing on any current operational experience – this helps to clarify current user needs. This is an in-depth, holistic discussion focusing on everyday and safety critical elements of a task. This helps to establish the 'as-is', from which existing assumptions and needs can be challenged to ensure 'new' is not simply a replication of 'current', and to identify issues or efficiencies in the current design.
    - Breaking the task(s) into sections whilst focusing on usability, performance and safety. This helped provide a semi-structured approach, giving structure and focus without limiting discussion and allowing other elements to be added throughout the session.
    - Mapping of the task goals, equipment and personnel interactions to the current design to identify potential issues.
- Reviewing several scenarios associated with the topic area, such as normal, degraded and emergency. These were established in advance where possible, however, it is necessary to retain some flexibility to include emergent scenarios during the session.

- Utilising ‘swim lanes’ to visually capture information on the different stages of the process being discussed for each scenario (see Figure 2).

There are often procedures or a task analysis which can describe the task and could be used when running the DITLO, however, it was important to not be too prescriptive in presenting the task steps to allow challenge of current, accepted norms and processes, and to limit assumptions being made. A semi-structured approach provided sufficient prompts to keep the session focused, but without constraining or limiting discussions. This helped ensure users discussed their task steps in greater detail and included the nuances of their working day that are unlikely to be captured by procedure.

Data Capture

A benefit of the DITLO process is the huge amount of rich detail generated. To ensure no data was lost and to aid facilitation and fluid discussion, data was captured in three ways during the sessions:

1. *Sticky notes capturing key information from the group presented in swim lanes* – visual representation of the tasks being discussed and the data captured within the room ensured delegates remained focused. Sticky notes were used to write down key information and presented within the relevant swim lane section of the task being discussed (see Figure 2).
2. *Detailed notes* – A scribe was also utilised during the DITLO to ensure all data were captured from discussions within the session. The scribe also captured actions if further information was required, or decisions needed to be made outside of the DITLO.
3. *The car park* – A ‘car park’ space was provided where off-goal topics, important to the stakeholders but not necessarily to the goal of the DITLO, were captured. This ensured that stakeholders felt their needs were noted and not dismissed, whilst allowing the workshop to progress and not sidetracked.

Output

A technical note was produced which summarised the key findings of the DITLO including actions, issues, risk, assumptions, observations and dependencies. The elements captured were not limited to HF considerations and provided insight and input across disciplines, promoting further engagement and integration between parties. Where HF specific aspects were captured, the associated documentation was updated accordingly, such as the Human Factors Issues Log (HFIL) and Human Factors Requirements Register (HFRR), in addition to informing aspects of the Task and Error Analysis. Other teams were also able to advance their documentation, such as requirements, process flow diagrams, early hazard identification, etc. The in-depth operational experience and task information gained through the DITLOs also permitted the identification of safety critical tasks. The technical note was circulated to the DITLO stakeholders for comment and input and, once issued, the document promoted a ‘paper-trail’ of evidence demonstrating how the design had matured and communicating the ‘golden thread’ of change over time.

Outcome

Benefits

In addition to the benefits outlined previously in terms of eliciting user requirements, validating assumptions and improving the understanding of how designs will be operated, other benefits were also noted.

The organic workshop approach provided benefits in understanding of the differences between what a user is assumed to do according to procedure, versus what they actually do and why there may be a difference. It also engaged end users in thinking about the reality of their day, rather than passively agreeing or not engaging in depth with the task steps presented. This delivered a detailed, real world understanding of the task, equipment and physical and personnel interactions. When mapped to the design, it highlighted gaps or inefficiencies or confirmed where the design philosophy worked.

Utilising non-technical skills such as clear communication and providing a collaborative workspace for conducting the workshop was fundamental to ensuring success and optimal output from the sessions. Providing a space where attendees feel psychologically safe, and valued to both offer their opinion and challenge others, is where a HF understanding of human behaviour can add benefit.

In an increasingly online or hybrid working environment, project teams rarely meet and share information in person. The DITLOs therefore provided an opportunity for multi-discipline integration and discussion in both a formal and informal context (e.g. during coffee breaks, breakouts, etc.), providing opportunities to overcome barriers to communication. This allowed attendees to collaborate and share a common understanding and purpose, and to build future collaborative relationships.

The process also allowed other disciplines (e.g. architects, piping, security) to understand the importance of early user input and learning from experience (LFE), which ensures design outputs understand and support key safety activities. In addition, it highlighted the benefit of HF involvement in refining requirements, amending system level requirements and identifying design changes early on, minimising the likelihood of costly re-design at later stages. Multi-discipline attendance at the DITLOs has helped bring the requirements to life, by providing an understanding the operational context from the end-users. For instance, understanding the underlying purpose of the requirement can not only ensure requirements are specific, it can also indicate what success or satisfaction of the requirement may look like.

Challenges

The DITLOs have not been without their challenges. Whilst the benefits of DITLOs have now been recognised within the wider project team, resulting in the stakeholder invite list to each DITLO growing, it has required a significant amount of effort to bring some stakeholders on the journey. Furthermore, the collaboration with other disciplines is necessary, but the increase in stakeholders at the DITLOs means that not all can attend in person and online stakeholders are not always actively participating in discussion. Further work is required to encourage online members to contribute more, rather than passively monitoring the workshop.

Ensuring the correct stakeholders are at the DITLOs can be a challenge, especially with competing demands on their time from others. Therefore, ensuring those essential to the running of the DITLOs are identified and early invites are distributed can help, together with a clear brief on the scope and purpose of the DITLO.

An additional challenge following the DITLOs has been distributing the information into the wider project team. The most effective approach is still being established, and may develop further as the DITLOs progress into RIBA 4 and become part of the routine project activities. However, given the success of the DITLOs, and the benefits they have added to the project to date – in helping to identify where design changes are required – it is likely that it will become easier to disseminate DITLO outputs.

Learning From Experience

Feedback forms are issued at the end of each DITLO to explore how useful the DITLO was to each stakeholder; the extent to which they were able to engage with the session; and what, if any, changes would they like implemented in future running of DITLOs. This allows continuous improvement on the administration, preparatory activities, facilitation, and running of blended events, to improve future DITLOs.

Next Steps

With each DITLO conducted, our HF team are learning about what works and what should be amended going forward. With the continuing success of the DITLOs supporting both HF activities and the wider disciplines within the project, it is likely that the DITLOs will continue to be run throughout RIBA 4, and across other projects due the valuable operational context provided. The DITLO method, although initially novel and looked upon with a certain amount of scepticism and fear, is now seen as an integral method adding value to project processes. It permits logical movement through both user tasks and activities, mapping them onto the desk and testing whether the design compliments both the user tasks and fulfils the stated purpose of the building.

Reference

Bye, R, (2017). DITLO: Finding the Value in the Everyday. Sixth International Human Factors Rail Conference, 6-9th November 2017, London.

Human Factors & AI In Nuclear: Regulatory Consensus As the Trust Bucket

Gorby Jandu

International Ergonomics Association

SUMMARY

In 2024, both the inaugural UK AI awards and Nobel Peace Prize for Physics in AI took place. This further intensified the mainstream publicity given to AI both in human automation and machine self-learning. In contrast, the uptake of AI in more industrial and specialist settings has been noticeably measured, especially away from simple automation. For example, in the nuclear industry, Human and Organisational Factors (HOF) have yet to benefit from AI en masse. This is at great odds with the potential that AI presents in mitigating nuclear harm, itself the main concern when scaling nuclear power plants. However, change is afoot.

In the last two years, important international nuclear regulatory guidelines for adopting AI in nuclear HOF have been published. This is a major step change for what is a decidedly risk-averse industry as it encourages the development of AI. But the message remains clear. AI must engender 'top-down' trust in the technology's deterministic predictability. Leading up to the published guidance, two world-leading 'sandbox' trials were conducted under regulatory auspices. Below, I discuss one in detail to show that nuclear can provide HOF practitioners with much-needed empirical data on the potential of AI. As, if nuclear can utilise AI then a vibrant case exists for other fields.

KEYWORDS

Nuclear, AI, Regulations

Introduction

Since the 'Turing Test' of 1950 intelligent technology has been reshaping the human-machine relationship. Despite this test being as old as nuclear technology itself, nuclear Human Factors practitioners have yet to adopt AI without circumspect. This is in part due to rigorous regulatory regimes that exist to mitigate the risks a nuclear facility poses to environment and societies. Key to this mitigation is the crucial role of the human. For example, the Level 7 accident at Chernobyl in 1986 was worsened by safety measures being ignored. Due to such accidents, the Atomic Act of 1946 established international regulations and co-operation that have only strengthened through international political order changes.

Another reason for the slower adoption of AI has been what Grote (2023) refers to as the 'reticence' of HOF practitioners to use newer technology. This has been due to the lack of clarity of how empirical evidence is produced leading up to classical HOF analysis. However, changes on both the transparency and the deployment of AI devices in nuclear is taking place. Its success has the potential to significantly change how HOF practitioners can support human welfare and performance outcomes far beyond nuclear surrounds.

Presenting the work as an independent academic, I briefly list opportunities for AI in nuclear HOF before presenting the challenges. This is much as Kirwan (2024) describes for the aviation industry.

Next, I foreground two case studies the UK regulator, the Office for Nuclear Regulation (ONR), is doing to encourage research into AI. Of these, I detail the HOF critique on one that has led to improved guidance by the regulator in adopting AI in nuclear. I argue that this dialogical process between HOF experts and the regulator has created a 'trust bucket' in the veracity of AI from design to deployment and beyond. This offers a signpost that AI and HOF can benefit human welfare and performance. The business-as-usual deployment of the robotic AI product (RrOBO) at Sellafield Ltd developed during these trials evidences this.

In all this, the measured and symbiotic nature of AI adoption by the regulator has been key to creating trust in this newer technology. A 'top-down' approach to AI development offers us a more classical analytical route over that seen in consumer electronics. This is important in nuclear as the consequences of an unpredictable event is morally and commercially critical. Hence, trust in this technology will be earned in drops as it can be lost in buckets.

This research is independent of the ONR and Sellafield Ltd, they remained informed parties.

Humans and Machines in Nuclear

Thorough License Conditions underpin the role of the human within the complex nuclear sociotechnical control system. As these are built on decades-long learning of human capabilities and vulnerabilities, they do not currently exist for AI Therefore, theory rather than practice has been the main beneficiary of research. Especially when discussing nuclear controls that ensure the safe, ethical, and beneficial use of AI in nuclear HOF. It should be noted that this is very much the status quo in outside of nuclear. For instance, as Salmon et al (2024) show for unmanned aerial vehicles.

All the same, the ONR states that AI in nuclear has clear benefits in safety, security, and safeguarding. Examples include the analysis of large volumes of data to better manage risks and improve efficiency; to accomplish tasks in hazardous areas in order to decrease the risk to workers; and potentially reduce error (ONR, 5th Sept 2024). In more specific research, Abdulrahman et al. (2024) have demonstrated that AI software can reduce the frequency of license event reports as correlated both to the number of vacant staff jobs and the ratio of contract employees to regular employees. Meenu at al. (2022) discuss this specifically for human factor errors in US nuclear power plants. So, in theory, there are significant gains to adopting AI in nuclear HOF. But only if safety and security is consistently demonstrable. This is why the role of the regulator is so vital in encouraging the safe testing of newer technology.

UK Regulator / Licensee Case Studies

In 2024, the ONR published two papers on AI, itself an uncommon occurrence in this industry. The second of these was notable as it set out considerations for the use of AI in HOF based on the 'sandbox' testing of two pre-deployment projects in 2023 (ONR, 8th Nov. 2023). One on structural integrity and the other on handling radioactive material. To exhibit the value of AI and HOF, I detail the latter as it involved human-retained automation of a previously labour-intensive human-machine interface glovebox function. This is used to handle highly radioactive material at Sellafield in difficult spatial and atmospheric conditions. Following observations and analysis by an industry-representative team, including HOF experts, it was concluded that assessed numerical claims on reliability could not be placed on the AI component. However, there was enough integrity in the trial process to encourage continued testing, resulting in deployment in 2024/2025.

As a result, the research and 'business as usual' use of AI in nuclear HOF continues apace at Sellafield Ltd. This has resulted in removing significant numbers of people from harm's way; a faster and more remote way of decommissioning hard to reach areas and finally; overall hazard and risk reduction in working conditions (ONR, 8th Nov. 2024). This is much the same as Sellafield

Ltd.'s everyday use of Spot the mobile robot 'dog' which was only deployed after trials successfully satisfied the strict License Condition 22: Modification or Experiment on Existing Plant.

The RrOBO Trial: Risk-reduction Of glovebox Operations Trial at Sellafield Ltd.

The ONR workshopped and tested the use of AI in a process where highly radioactive material is processed in a challenging operational environment. The objective for HOF practitioners here has always been to minimise human exposure to time spent with radioactive waste. The goal in inviting HOF experts was to substantiate safety claims that moving the human to a supervisory rather than operator mode would decrease the level of threat and incident.

The Risk-reduction Of glovebox Operations (RrOBO) mechanical system details are as follows: the baseline system for active deployment consists of a Kinova robotic arm protected by a containment sleeve and equipped with a parallel jaw gripper as the end effector. It is compatible with future tooling and mobile stands. Cameras are inbuilt to allow for remote viewing and remote operator control of the robotic arm via a haptic controller with force feedback.

The supply chain demonstrations included fitting the containment sleeve, inserting the robotic arm into a glovebox and describing the selection of additional tools and capabilities that would become available through COTS (Commercial-Off-The-Shelf) items. These tools and capabilities include applying decontamination gel, swabbing, LiDAR (Light Detection and Ranging) laser scans, radiometric sensors, and many more. (Source: Sellafield Annual Research and Development Review, p. 26). Below in Fig. 1 are some images from Sellafield Ltd.

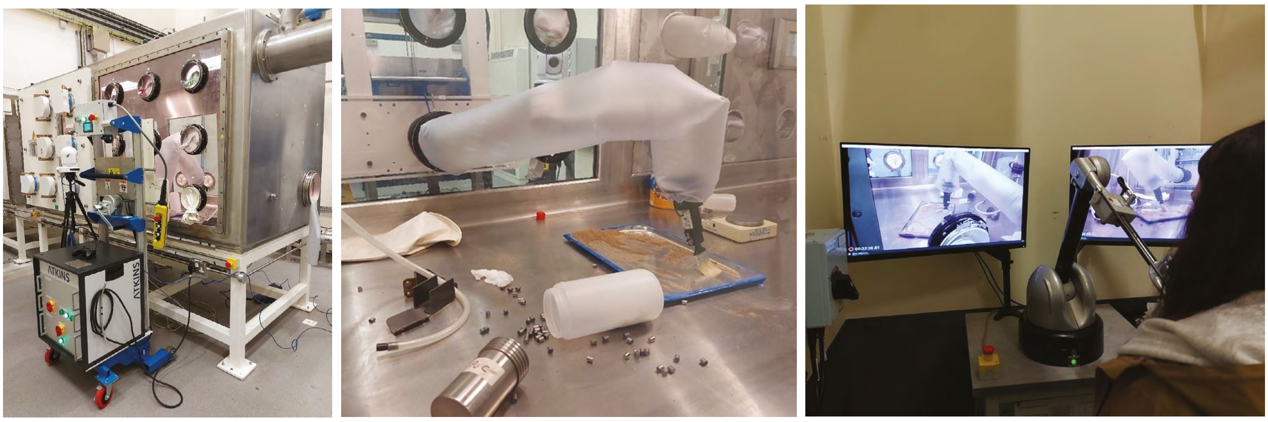


Fig. 1: Demonstration of the Kinova robotic arm in a nuclear glovebox; inactive trials of the Kinova robotic arm in a nuclear glovebox; Kinova robotic arm control room. (Source: Sellafield Annual Research and Development Review, p. 26).

Table 1 below summarises some of the topics and outcomes of the pre-trial workshop as commented by the assembled team, including HOF practitioners and researchers. A clear pattern of unpredictability emerges.

Table 1: Pilot of a regulatory sandbox on Artificial Intelligence in the nuclear sector

Topics	Outcomes
The possibility of placing a numerical claim on the AI component itself;	The group did not consider it possible at present and attributing reliability numbers to an AI system could drive the wrong behaviours.
Design of the whole system, including wraparounds and other systems to support the safety, security and environmental case;	It is likely to be difficult to substantiate AI software due to their complexity, inaccessibility to code to understanding the machine learning algorithms.
Human/AI coworking, managing handovers from AI to humans, recovery from faults and defining a safe state; and	The group noted that 'the human' cannot be removed from the system altogether.
Modifications and future behaviour, maintenance procedures and through-life behaviours.	Subtle changes in the AI operational environment have the potential to change the performance significantly. The sandboxing revealed that there may be a gap in the availability of good practice for the examination, maintenance, inspection and testing of AI systems.
Conclusion	
It was recommended that AI algorithms should be locked down and arrangements put in place to control modifications. Deep, continuously learning systems could be used as a non-safety aid but are unlikely to be justifiable for use in any safe operation. It was considered that self-learning systems may challenge current guidance for Licence Conditions.	

Source: Regulators' Pioneer Fund (Department for Science, Innovation and Technology): Pilot of a regulatory sandbox on artificial intelligence in the nuclear sector.

Analysis and Commentary

It is clear from the above that HOF experts were unanimous in being unable to substantiate safety claims on the deployment of AI robotics in this particular setting. This is not uncommon for this discipline and for the state of AI in other critical safety industries. However, since the sandbox trial of 2023, the RrOBO devices are now scheduled for active deployment into several gloveboxes on site. It was approved at the Remediation portfolio board to be delivered in 2024/25 (Sellafield Annual Research and Development Review, p. 26). It is notable that significant contribution has been made by HOF practitioners in this achievement toward reducing the risks to humans whilst improving productivity.

As a simplistic example of HOF contribution, in Table 2 is a comparison of the original definition of this AI application to a robotic glovebox task and then the revised problem/opportunity statement. The imprint of HOF conceptualisation and pre-design expertise is clear.

Table 2: Comparison of approach statements

Original	Revised
AI for real-time application to inform operations to optimise robotic movements.	The robotic glovebox will process a highly radioactive material that needs to be processed in a challenging operational environment where minimising human exposure is critical/important. This is currently done manually by technicians and is very labour intensive. Main hazards are leakage into environment (generating large amounts of radioactive waste) and danger to the operator. There are multiple

	possible deployment modes –from full mission-based automation, to acting as a supervisor for a human controlling the robot. The focus of this work will be on the case both for the supervisory robot (where the arms are teleoperated by a human, but the AI system provides haptic feedback to prevent collisions or other dangerous behaviour) and autonomous mission-based modes (with a human supervising the operation remotely). The arms could be handling tools during operation.

The above formed the pre-cursor objective before the sandbox trials were undertaken and the commentary in Table 1 developed. As such, it would not have been possible to develop a critique on the safety claims without a thorough problem/opportunity statement. This itself would not have been possible without the input of HOF academics and practitioners. So, whilst HOF might show some reticence towards the role of AI, in fact development in the nuclear industry show that the measured and symbiotic development of AI has only been made possible through HOF practices and research.

Further to this, what ought to give HOF, robotics and AI encouragement is the measured and collaborative process undertaken by the UK regulator. Also, the list of collaborators and contributors is comprehensive enough to account for expertise amongst academics, commercial interests to execute, international regulatory bodies and nuclear power plant licensees.

Conclusion

AI in nuclear HOF is undergoing significant research globally, with the UK 'sandbox' trial the first of its kind anywhere in the world. This, coupled with significant literature output from collaborating international regulators shows that the role of the human in the nuclear industry could be set to change. Whilst HOF has been slower to adopt AI into traditional assessment methodology, the collaborative and measure pace of change in nuclear seems to suit the way HOF has approached safety claims and human/machine task allocation. In many ways, both nuclear workers and HOF practitioners appear to understand that it is this calculated change of pace that will create the all-important trust bucket for AI predictability. In nuclear, it is very well understood that, should it tip over it will empty very quickly.

References

Abdulrahman K., Abdulelah M. A., Rajasekaran S., Shanmugasundaram M., (2024), Human factors engineering simulated analysis in administrative, operational and maintenance loops of nuclear reactor control unit using artificial intelligence and machine learning techniques, *Heliyon*, Volume 10, Issue 10, e30866, ISSN 2405-8440, https://doi.org/10.1016/j.heliyon.2024.e30866. (https://www.sciencedirect.com/science/article/pii/S240584402406897X)

Kirwan, B. (2024) The Impact of Artificial Intelligence on Future Aviation Safety Culture. *Future Transp.* 2024, 4, 349–379. https://doi.org/10.3390/futuretransp4020018

Meenu Sethu, Bhavya Kotla, Darrell Russell, Mahboubeh Madadi, Nesar Ahmed Titu, Jamie Baalis Coble, Ronald L. Boring, Klaus Blache, Vivek Agarwal, Vaibhav Yadav & Anahita Khojandi (2023) Application of Artificial Intelligence in Detection and Mitigation of Human Factor Errors in Nuclear Power Plants: A Review, *Nuclear Technology*, 209:3, 276-294, DOI: 10.1080/00295450.2022.2067461

Paul M. Salmon, Scott McLean, Tony Carden, Brandon J. King, Jason Thompson, Chris Baber, Neville A. Stanton, Gemma J.M. Read, (2024) When tomorrow comes: A prospective risk assessment of a future artificial general intelligence-based uncrewed combat aerial vehicle system, *Applied Ergonomics,* Volume 117, 2024, 104245, ISSN 0003-6870, https://doi.org/10.1016/j.apergo.2024.104245. (https://www.sciencedirect.com/science/article/pii/S000368702400022X)

Office for Nuclear Regulation. (2024, 5th September) *New paper's international considerations for AI in the nuclear sector*. Technical Directorate. Retrieved 11th November 2024, from https://www.onr.org.uk/news/all-news/2024/09/new-paper-shares-international-principles-for-regulating-ai-in-the-nuclear-sector/

Office for Nuclear Regulation. (2023, 8th November) *Outcomes of nuclear AI regulatory sandbox pilot published.* Technical Directorate. Retrieved 5th November 2024, from, https://www.onr.org.uk/news/all-news/2023/11/outcomes-of-nuclear-ai-regulatory-sandbox-pilot-published/

Testing HF Requirements to Optimise Human Performance

Chris Heath & Ewan Povall

RPS Group, UK

SUMMARY

This paper will cover the approach used to integrate Human Factors (HF) requirements into project testing and commissioning phases of a major nuclear new build project as well as key insights to reliably test the design and tasks involved to optimise human performance. This includes discussion on the importance of deriving good HF requirements, testing these requirements effectively and delves into the HF Verification and Validation (V&V) techniques involved to successfully progress the design project from detailed design into testing and commissioning phases. With the overarching goal to produce a safe and operable facility.

KEYWORDS

Verification, Validation, Requirements, Testing

Introduction

Within high hazard industries a failure in human performance has the potential to lead to significant consequences to the operators and public. People regularly talk about human performance in regard to the "doing" or "sharp end" of a task. However, often human performance is influenced long before we ever start a task – when defining the required role of the operator, throughout the iterative design process, and validating the design through testing. In safety critical operations we cannot just purely rely on people to do the right thing at the right time – we have to provide the environment for them to succeed and part of that is through the task design and design of the equipment and systems they use. Ensuring that the system is designed so that the right thing to do is the easiest thing to do, and then confirming those tasks can be performed as intended.

It is important therefore, to have HF requirements which are well derived and managed, to guide the iterative process from concept to detailed design. Beyond design phases it is crucial to continue building confidence throughout testing that the operator role is adequately supported by the design, and where necessary identify operability issues and adapt the design to optimise human performance.

To ensure that the design meets HF requirements (verification) and for the design to achieve its intended purpose (validation), it is vital to gain reliable evidence by having set testing criteria for each requirement and fully integrate this criterion into testing plans. Embedding criteria into testing plans to reliably verify and then validate the operability of the design is key for ensuring it is optimised to support human performance during operations, maintenance and recovery tasks.

Why test the role of the operator throughout testing and commissioning phases?

- **Achievability of Tasks** - Testing the role of the operator aims to provide confidence in the operational design via physical demonstration and testing that the required tasks can be achieved. Engineering design requirements help ensure that the equipment functions as expected and in a reliable manner, and equivalent to this, are HF design requirements which are essential to demonstrate that equipment is feasible to operate and maintain safely.

- **Drive and optimise design changes where required** – Where testing identifies operability issues with the design verification or where functional issues drive changes which impact operability, testing HF requirements and ensuring criteria is accurately tested and validated will ensure the final design successfully meets the requirements.
- **Project de-risking** - If HF requirements are not tested, the project significantly increases the amount of risk it carries into the operational design, as issues are identified late in the process or potentially go unrealised. In worst case scenarios this will lead to an unsafe and inoperable design placing an unacceptable and potentially unachievable amount of reliance on Human Performance.
- **Time, Process, and Cost Saving** - Early testing is the ideal opportunity to confirm that the manufactured design meets HF design requirements. This provides the ability for HF issues to be identified and resolved as early as possible when it is the most timely, simple, and cost effective to make the necessary changes to the design.
- **Regulatory Compliance** - The Office for Nuclear Regulation (ONR) and other relevant regulatory bodies have set conditions which the design of this facility is required to comply with. Validating HF requirements and recording evidence during testing and Commissioning phases is a vital step in demonstrating this.

How do we successfully test the role of the operator?

Derivation of optimised requirements

Having the project's set of HF requirements derived through both generic and specific methods provides the means to comprehensively verify and validate all possible design aspects of the system which impact human performance. The two methods are depicted as follows:

- **Generically**, from standards/good practice guides or,
- **Specifically**, from task analyses/HAZOPs/Safety Assessments (e.g., Human Based Safety Claims or Safety Critical Tasks), to support specific elements of the task.

This combination of derivation methods helps to form requirements which enable the operator to achieve all individual tasks and ultimately the task or system goal in a safe manner.

On the major nuclear new build project, HF requirements have been derived early within the project lifecycle and integrated with other engineering disciplines requirements. Through an integrated approach, we were able to ensure:

- HF requirements were worded to enable the design to demonstrate how it would be met, while not restricting the means by which it should be achieved (see following section for more detail).
- HF requirements were treated with equal importance to engineering requirements.
- Requirements were individually identified, tracked and linked to the safety case and engineering safeguards.
- Duplication was minimised and prevented where possible.
- Requirements were assigned to the right people for implementation.
- Communication of requirements is as easy to understand as possible.
- Eased the process to integrate design requirements into testing phases through developing clear test and success criteria.

In addition, should existing requirements be modified or new requirements derived as and when the need arises (e.g., due to changes in operational philosophy, new human-based safety claims from the safety assessments, design changes, etc.,), this is easily managed using a HF V&V Plan.

Characteristics of optimised requirements include:

- **Worthwhile:** Is the requirement genuinely achieving something in the design or supporting an operator task in some way?
- **Targeted to the audience**: Does the requirement make sense to the individual responsible for implementing it?
- **Genuinely required:** Is it a genuine requirement or guidance? Guidance should be managed separately to support designers in their work.
- **Solution agnostic**: Focus on what the equipment/system needs to do or provide, not how it will be achieved, this helps to ensure relevance even if the design changes and allows designers to identify optimal solutions.
- **Tailored:** Is it relevant to the specific operations and human performance standard required?
- **Measurable:** Are there clear ways of verifying and validating the requirement?
- **Balanced:** Is the requirement trying to do too much or being hyper focused when it could be more generic. Or perhaps the opposite is true, is the requirement too generic which may cause it to become difficult to understand the parameters to fully validate it?

Develop applicable testing criteria

From well-defined requirements, HF testing criteria has been developed to be appropriate for each stage of testing based on the planned scope, availability of equipment and elements of the system, and the possible conditions of testing. In collaboration with relevant stakeholders including Engineering, Pre-Operations, Maintenance, and Commissioning teams; how and at what stage testing criteria can be reliably tested has been confirmed. HF design requirements have been added to test plans where applicable, with task conditions accurately replicated as early and as close to the true operating conditions as possible. Successfully replicating conditions will maximise the level of validity in the evidence produced by the test results.

Unlike purely functional Engineering requirements, standard pass/fail criteria is often not suitable to successfully test the majority of HF requirements which focus on assessing the operability of the design. A range of testing criterion types have been identified to be required to effectively test the full range of HF requirements selected for testing. An overview of the testing criterion types implemented into the project's testing scope includes:

- **Manufacturing Confirmation:** Requirements which require confirmation that the manufactured design is accurate to the intended design e.g. confirm the Glove port height is within the ergonomic range; a pass/fail criterion within a set of pre-defined limits.
- **Trial Confirmation:** Other requirements may gain sufficient V&V evidence from testing through appropriately designed test steps. These suitably test the equipment and tasks involved against the criteria to gain the required confidence for that testing stage. An example of this could be a step confirming whether the package has successfully been cut, which is directly linked to an HF requirement concerning the achievability of this action.
- **Operational & Witnessing Subjective Feedback:** For HF requirements that are more operability focused, the most significant method to gain valuable validation evidence is through conducting the tasks under closely replicated conditions and recording live subjective qualitative feedback from the test operators and the witnessing stakeholders present for the test (including HF or Pre-Operations personnel if possible). For testing these

requirements accurately, defined testing conditions are crucial to gain reliable validation evidence.

The level of validation possible to gain will increase throughout the testing and commissioning phases as elements of the built design connects, eventually within the operational environment, at which point the conditions will be fully replicated. At active commissioning the materials to be processed within the facility will be involved as the final step to gain full confidence in the operational process and ability to successfully manage the hazards involved.

It is worth stating at this point that the majority of verification activities have been carried out within the detailed design phase and are captured within the HF V&V plan prior to testing. Verification evidence is necessary to ensure the majority of the design is confirmed prior to initiating the manufacture stage. Subsequent testing phases then focus primarily on the validation of the manufactured design with the aim to gain validation evidence. However, depending on the maturity of the design related to the requirement being tested, verification activities may still be important to complete via early Factory Acceptance Testing (FAT) or even after during Integrated Works Testing (IWT) if required. This is particularly relevant where design changes have occurred.

Method used to effectively integrate HF requirements into testing

To successfully embed HF requirements within testing plans, testing criteria was developed which is realistic considering the availability of equipment and replicability of tasks to the true operational conditions. The HF integration method was developed for this project as follows:

- **Step 1: Selection of HF requirements to test** – All design requirements contained within the HF V&V plan were reviewed to identify those feasible to test at each stage. This screening activity was made in collaboration with the Engineering, Commissioning, Pre-Operations and Maintenance teams and is dependent on the availability and possible setup of equipment, and scope of testing at each testing and commissioning phase.
- **Step 2: Derive test and success criteria** – Test and success criteria was derived for each requirement with any special conditions required to achieve the level of verification and validation expected.
- **Step 3: Integrate HF design requirements & criteria within testing documentation** – In collaboration with test authors, the test steps which were applicable for testing each requirement were identified. To track and manage requirements, unique HF requirement IDs were assigned to individual steps. HF judgement was required to assess whether the inclusion of HF requirements criteria was acceptable. In some instances HF have required further testing or for a completely separate test performed to adequately test the criteria at this phase. This is the case for HF requirements identified to have important tasks associated and/or help to underpin key Operational Preventative Measures (OPMs) critical to maintaining safety.
- **Step 4: HF review of testing procedures** – The HF review of test procedures has helped to ensure criteria has been implemented consistently, testing conditions and pre-requisites are well defined, and the correct input has been provided into the design of the test and individual steps. This input is key for HF requirements to be tested by the expected method and therefore, produce reliable testing results. HF are key stakeholders to the development of testing documents and in order to progress, HF approval is required.
- **Step 5: Tester and Witnesser recording of evidence during testing** – Accurate and reliable evidence recorded during testing. HF validation guidance to support the subjective feedback required of each HF requirement has been provided along with HF briefings to the witnessing team in order to record suitable evidence for HF requirements., HF form part of the Witnessing team present at selected tests to ensure a proportionate and

pragmatic level of involvement. HF have identified specific tests which require HF to attend judged on the importance of assessing the equipment and tasks involved.

Supporting HF Activities and Deliverables to Optimise Integration

In addition to integrating HF requirements and criteria into project testing documents following the above process. Further support has been provided to ensure the testing process fully assesses the design's operability. Supporting activities with stakeholder buy-in have been conducted to produce and implement the following documents:

- **Glovebox setup and equipment use document –** The purpose of this document was to agree and record the Glovebox setup, availability and use of equipment during Factory Acceptance Testing (FAT) and Integrated Works Testing (IWT) to ensure the HF requirements tested will replicate the as built/manufactured design as much as possible.
- **Glove port position assessment –** This document provides a checklist assessment template to evaluate the Glove port positions at the earliest possible stage in testing to confirm positions and make any necessary adjustments to optimise operability. Reach, force, and visibility aspects are assessed in context to the full range of tasks associated with the equipment and components to be interacted with through each port.
- **Guidance for Validation of HF Requirements –** This document provides guidance to the testing team on the method to successfully record validation evidence for HF design requirements being tested when HF are not in attendance.

Table 1 below is an example taken from the validation guidance document to help inform the testing team, including facilitators, operators, and witnessers of the questions and considerations when reach is a relevant factor during testing. The intent of this checklist is to provide an understandable set of considerations applicable, as a guide to record sufficient qualitative evidence when evaluating whether the as built design supports the operator's ability to reach and manipulate equipment within the Glovebox.

Table 1: Reach Validation Checklist

HF Requirement:	
V-HF-50: The design of gloveboxes shall ensure that equipment to manipulate is within the comfortable zone of reach for the 5th percentile female where practicable for normal, maintenance and emergency procedures.	
HF Criteria:	
Gain subjective feedback from test operators on any issues found to reach equipment and areas of the glovebox involved with the test step.	
Test Witnessing Questions/Considerations:	
1.	**Reach all Equipment:** Can the tester reach all of the equipment that they need to interact with? Identify the Glove Port and equipment with reach issues to be considered for adjustment.
2.	**Task Completion:** Can the tester complete the required tasks when having to reach the equipment?
3.	**Operation:** Can two hands be used to operate equipment where intended?
4.	**Additional Testers:** Did the tester need any support to perform any tasks (from another tester)? e.g. additional testers required to lift/lower/push/pull. Specify how many if additional required.

5.	**Other Users:** Do you think other users than the tester may experience issues reaching and/or completing the required tasks? i.e. Someone with different physical characteristics; larger or smaller in stature, with more or less arm reach, larger arm breadth etc. Specify tasks and/or specific parts of tasks which apply.
6.	**Body Positioning:** Does the tester adopt any awkward body and limb postures to complete the tasks? i.e., leaning, squatting, turning side on to allow more reach with one hand, moving body or arms in unnatural way to reach? Detail each relevant part of the task and what the issue is.
7.	**Accidental Operation:** Are any issues with accidental operation / touch anticipated due to reach issues experienced?
8.	**Replication of Task:** Tasks carried out as if on plant? E.g. Level of PPE used. Any operational impact of PPE use e.g., Does the use of PPE create any issues for reaching equipment?
9.	**Use of Aid/ Tooling:** Is the use of testing aids / tooling required to reach any component? If so, does the designed tooling help the tester reach all of the equipment that they need to interact with and complete the required tasks? Is there any additional tooling needed not previous expected? Any issues with Aid/Tooling used? i.e. hard to use the aid, time consuming, increases difficulty of other task elements etc.
10.	**Error Traps:** Are there any 'error traps' identified while performing the task?

Recording of Evidence and Approval of Results

Evidence is recorded within the test procedure document against each step and then signed-off by each required stakeholder. For identified tests an HF representative will be required to confirm and approve the evidence recorded. Where applicable HF subjective feedback is recorded against the test step, and additional notes and assessments will complete the recorded evidence tied into each completed test. Where HF related feedback has the potential to impact the design and operations, formal observations are made, and more significantly any major issues are captured as a formal Test Fault Observation (TFO).

Following completion of the test, a test results approval process is followed, and HF approval is required as a key stakeholder. Observations and TFOs are reviewed, and where possible resolutions are progressed and re-tested prior to the start of the subsequent testing phase. HF review completed test documents, and where HF did not attend, this review provides a vital hold-point to ensure confidence is gained in the validity of the results produced.

Once the process is completed test documents are approved, HF then update the relevant testing sections of the HF V&V plan with evidence including references, and as necessary also update the testing criteria of future testing phases with any changes as a result of the completed tests.

Key insights and challenges to HFI into testing

A key element to integrate HF testing criteria successfully is to inform the setup and use of the equipment to be tested, achieving as close to operating conditions as possible. To achieve this, true reach distances and physical operability conditions are simulated through attaching a replicated Glove Port mock-up to the temporary testing windows of the Gloveboxes. A range of additional tooling and testing aids have been identified for use during selected tests including operator steps/platforms, as well as Personal Protective Equipment and Respiratory Protective Equipment such as glove combinations and respirators.

Testing as planned, as performed, and as recorded can have greater differences than expected. Many variables impact testing, and this highlights the importance of providing accurate and well detailed pre-conditions for testing and where differences arise, for these to be captured as part of testing to ensure recorded results are reliable. The importance of test wording is a crucial element of this, to ensure that the order, technique and equipment position or state from which tasks are performed are not open to interpretation by the testing facilitator and test operators.

Highly important is to gain multi-discipline buy-in early to the testing criteria demands which are necessary to successfully test HF requirements. If other key stakeholders understand these reasons from the outset, the HF input to testing will result in a much greater impact to validating the role of the operator. Having sought early engagement, the experience of Pre-Operations, Engineering, Maintenance, and Commissioning teams is utilised to confidently gain validation evidence and identify any additional HF related issues.

We have found that although only the functional elements may be tested first, it is natural for observations concerning the operability of equipment to be made early (prior to planned operability testing). To make the most of early opportunities and assess HF requirements throughout the testing process, using operational experience such as the Pre-Operations team can help ensure that HF requirements are assigned and considered to earlier functional test steps which may benefit from their inclusion even if operability is not directly tested as part of the procedure.

Collaboratively Integrating Inclusive Design into Very High-Speed Rolling Stock Development

Aoife Finneran, Simon Cran & Melanie Tse

Alstom

SUMMARY

This paper describes a collaborative working arrangement between a very high-speed train Purchaser (HS2 Ltd) and a Train Manufacturer and Maintainer (Hitachi Alstom High Speed, HAH-S) to ensure deployment of inclusive design principles in the design of new trains.

KEYWORDS

Inclusive Design, Collaborative Design, Rolling Stock

Introduction

The Train Manufacturer and Maintainer (TMM), HAH-S, has been given the contract to design, build, and maintain very high-speed trains for the Purchaser, HS2 Ltd. HS2 Ltd has set strategic goals to achieve new standards in customer experience placing people at the heart of design and therefore requires its suppliers to take an inclusive design approach. Key lessons learnt as part of the HS2 project are more generally captured on the HS2 Learning Legacy website. A key priority for the TMM is ensuring that the product integrates inclusive design principles throughout the design process for the passenger environment. Rail transport is a complex socio-technical, safety-critical system and hundreds of stakeholders are involved in the design, development, manufacture, and testing of new trains. Rolling stock design and manufacture exists within a rigid mandatory standards framework, often with multiple competing requirements that can impact passenger experience (Read, 2019) and (Aya Bayramova, 2023). This paper describes how an inclusive design process was developed and deployed against this complex backdrop.

Train Design Process

Simplified in Figure 1, the delivery of rolling stock projects is complex and HS2 is no exception. The process involves multiple specialisms and engineering disciplines where teams can be located globally in large multi-nationals or within contractor and supplier organisations. The delivery of significant milestones and project deliverables and their approval is managed over three design phases (Concept, Preliminary, and Detailed) using the Engineering Design Review process. During these design phases the train progressively matures until detailed design freeze is reached, and no further changes are made without an additional change request process. During the design phases there are meetings, interactions and review of documentation with the HS2 Ltd and their shadow operator West Coast Partnership Development (WCPD). These are managed using the HS2 Ltd Stakeholder Management process until final testing is complete and handover of the rolling stock to HS2 Ltd for operation on their network.

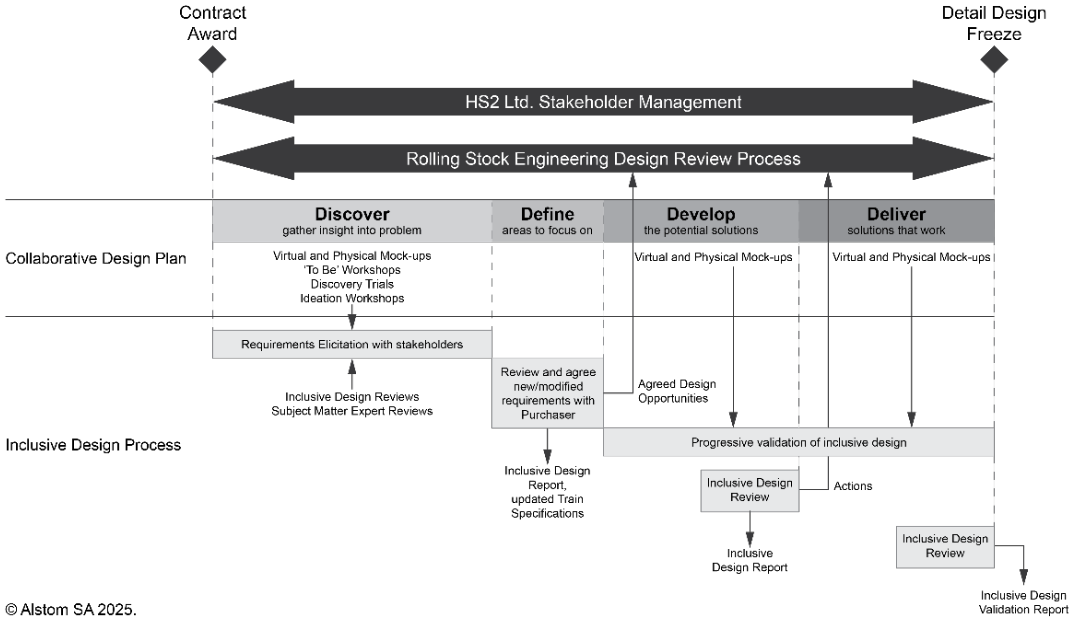


Figure 1: Inclusive Design Process within the Very High-Speed Rolling Stock Delivery Plan

Collaborative Design Process

During the Tender phase and in advance of the formal project award for the HS2 Rolling Stock, several design challenges and scoping activities were undertaken by HAH-S to define appropriate collaborative design activities over the life cycle of the project and means by which to elicit input from stakeholders and end users (Figure 2). This Collaborative Design process forms part of the TMM's commitment to work collaboratively with HS2 Ltd and associated stakeholders. The Collaborative Design Plan (CDP) describing the process was agreed with HS2 Ltd following Contract Award. The Collaborative Design process aligns with the Double Diamond model developed by the UK Design Council (Drew) and aims to leverage the experience and capabilities of the TMM and their partners. Collaborative Design activities included: To Be Workshops, Discovery Trials, Inclusive Design reviews, and other Human Factors (HF) workshops and trials. Each activity engaged different stakeholders including passengers, crew, operational staff, and the HS2 Ltd design team to ensure a holistic approach to the train design and development. These reviews used wooden mock-ups or Virtual Reality (VR) representations of the train design and considered Day in the Life of (DITLO) task scenarios and task analysis prepared by the HAH-S Operability and HF teams. The review process was iterative with each activity feeding into the next. HAH-S also supported and considered parallel design activities undertaken by HS2 Ltd the Built Environment Accessibility Panel (BEAP), and their delivery partner WCPD. The Discover and Define phases used the as-sold design as a baseline, which the initial wooden and VR mock-ups represented. As illustrated in Figure 1 the collaborative design process aligned with and fed into the Rolling Stock Engineering Design Review Process. The Rolling Stock HS2 Ltd Stakeholder Management reviews considered the findings from the collaborative design activities, in terms of technical and financial feasibility in line with HS2 Ltd requirements and specifications.

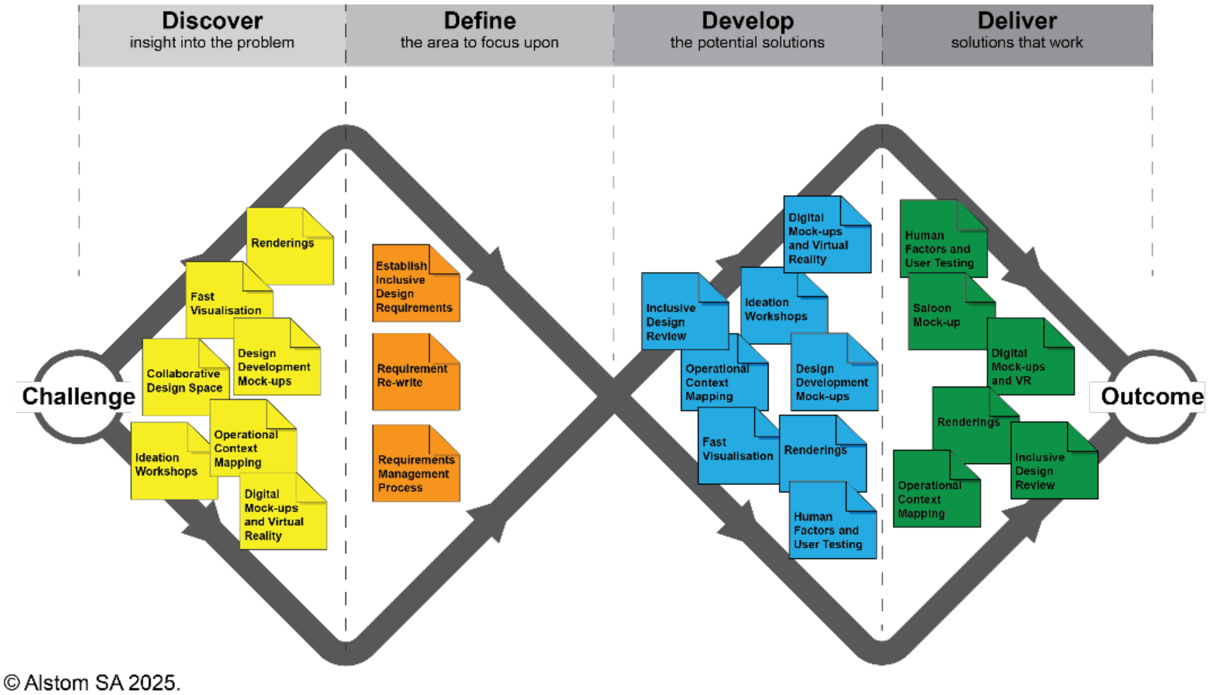


Figure 2: Collaborative Design Activities

Inclusive Design Process

Between Contract Award and rolling stock Detail Design Freeze, the Inclusive Design activities form part of a wider, iterative Collaborative Design process (Figure 1) following the principles of the UK Design Council's Double Diamond framework (Figure 2). The Inclusive Design Process is complementary to the Engineering Design Review process for the trains. Activities were aligned with each of the design phases of the Double Diamond and were reviewed iteratively with relevant stakeholders to ensure that they considered the contractual and regulatory demands of the project. Several co-design activities were undertaken with representative end-users from passenger and staff groups using a variety of tools including VR and physical mock-ups of the train interior design. These activities were iterative with potential design changes considered and reviewed in subsequent phases. The Inclusive Design work package was delivered by Alstom's HF team and independently audited and approved by an external auditor Instituto de Biomecánica de Valencia (IBV) who both have extensive experience in the delivery of inclusive design projects.

Inclusive Design Review

The British Standards Institute (BS700-6:2005) defines inclusive design as "The design of mainstream products and/or services that are accessible to, and usable by, as many people as reasonably possible…. without the need for special adaption or specialised design." As defined in the CDP, the Inclusive Design Review complements other workshops and user testing with representative end-users by collecting feedback from Subject Matter Experts (SME) in Inclusive Design. Taken as a whole, the series of inclusive design activities aim to provide progressive insight into the impact of design choices on the inclusivity of the train, utilising VR, photorealistic images, and wooden mock-ups of the train interior. A semi-structured questionnaire was used by HF specialists to guide the workshops as per the methodology in (Afacan, 2009). The following areas were assessed as part of the define phase review, the Platform Train Interface (PTI) and Exterior Doors, Toilet Amenities, saloon, and café/kiosk areas, bike storage, buggy storage, and passenger facing catering areas.

Results from the inclusive design sessions were transcribed and analysed (using thematic analysis) by the HF team. Potential inclusive design solutions, as described earlier were reviewed for technical and financial feasibility by HS2 Ltd and HAH-S engineering teams against the as-bid train solution. The implementation of agreed inclusive design changes is tracked using the Design Review Open Items List (DROIL) where system owners give regular updates of design development progress.

Results

This section describes some inclusive design solutions considered as part of the inclusive design review process.

Grab Poles

HS2 Ltd requirements are distributed to the TMM via the Train Technical Specification (TTS). The TTS requirement for the minimum exterior door horizontal clearway is 900 mm, when the door is open. HS2 Ltd requirement is 100 mm more than the minimum mandatory requirement from National Technical Specification Notice (NTSN) Persons of Reduced Mobility (PRM), a mandatory standard for accessibility. The additional 100 mm provides additional clearance for wheelchair user's hands, crutches or a walking frame. To achieve the required horizontal clearway, the grab rails were sunk into pockets either side of the exterior door maintain the 900mm throughway. The design was incorporated into the mock-up and tested with user groups as part of the Discovery Trials. Twenty-six participants took part in 6 workshops including users with mobility, hearing and visual impairments, and users who were neurodivergent. The user groups reported challenges with the design of the grab poles at the exterior door. Users noted that the edge of the pocket resulted in awkward wrist postures and, when standing on the platform in front of the moveable footstep, users needed to reach forward at the waist or bend at the waist to reach the grab poles, and this was an issue when alighting or boarding the train, as shown in (Figure 3). The findings were confirmed in the Inclusive Design Trials with three SMEs in Inclusive Design leading to a recommendation "To improve access for users, consider moving the grab handles in Y (towards the platform) and X direction, closer together" was given to HS2 Ltd.

Figure 3: Testing of grab poles in doorway

Figure 4: Testing of grab poles in doorway with updated mock-ups

Iterations of the grab pole design were developed by Alstom's Advance and Creative Design (A&CD) and Train Design team considering the system integration, mandatory standards, and accessibility constraints. Each design proposal was reviewed by the HF team utilising CATIA V5 and Human Builder (Figure 5). The chosen design was implemented into the mock-ups and tested positively with user groups and SMEs in a second round of assessment (Figure 4).

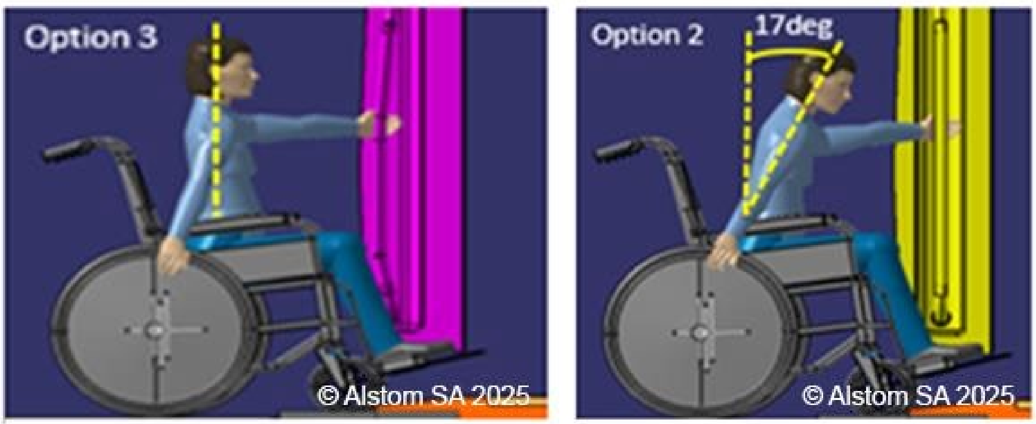


Figure 5: CATIA V5 and Human Builder assessments of grabpole accessibility

Bike Stowage

The bike stowage area in the as-bid design involved vertical racks requiring users to wheel their bike in horizontally, then lift and rotate it to store it as shown on the left in Figure 6. Most participants found this task difficult and was made worse if, one bike needed to be moved before stowing a second bike, especially if the first was not stowed correctly. The recommendation from user trials was to make bike storage horizontal.

Following discussions with HS2 Ltd a new layout was developed by Alstom's HF, A&CD and Train Design teams. The new layout, shown on the right in Figure 6 considered how users move through the train as the bike stowage area is also a throughway for passengers. The design team considered many different factors to ensure the layout provided sufficient grab poles whilst also allowing convenient and efficient bike stowage and removal. The HF team used the good practice and guidance in EN 16585-2 (BS16585-2:2017) and undertook investigations using CATIA V5 and Human Builder with the design team to optimise the grab pole positions for bike users and other users including those with accessibility needs. The revised layout provides horizontal storage of four bikes with 2 on each side of the aisle and each bike offset from its neighbour allowing bikes to be stowed and retrieved independently.

The new layout was presented in a mock-up to user representatives and SME and positively received by all participants. One participant asked whether "the angle of the grab pole inhibits smooth movement, the lower edge of the grab pole is clashing with the bike, can we shorten the grab pole?" The HF team highlighted this suggestion to HS2 Ltd but explained that the assumption was that bike users were able to successfully manoeuvre their bikes priority should be given to providing sufficient support for passengers walking thorough the saloon. HS2 Ltd agreed the assumption was reasonable, and no further changes were needed.

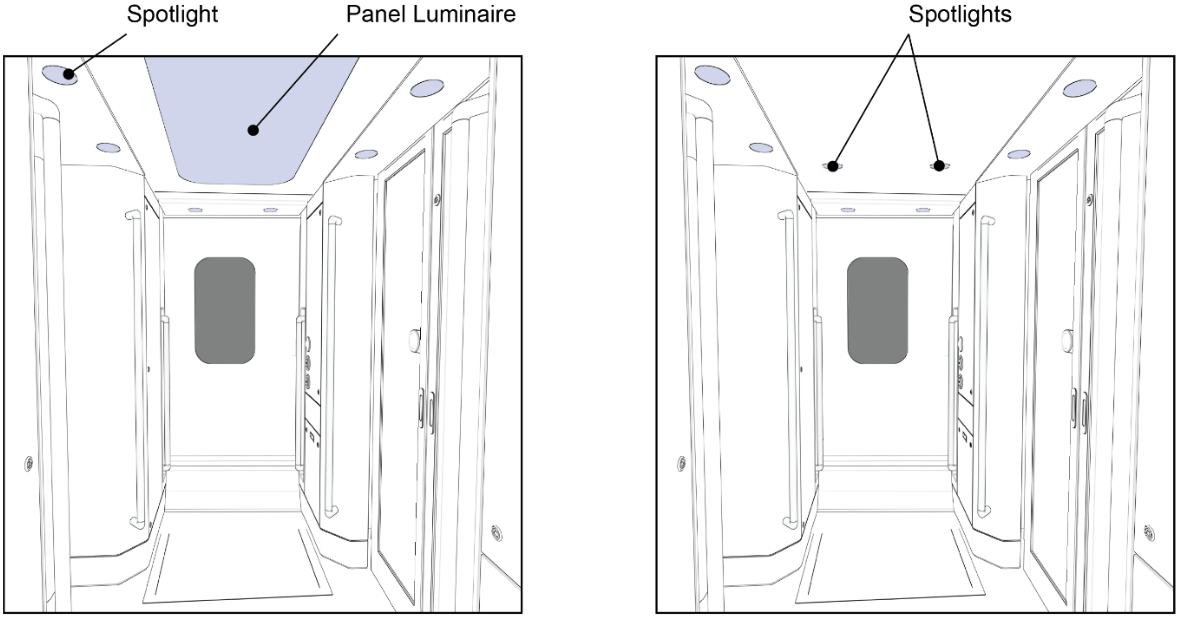
Figure 6: Vertical bike storage (Left) and horizontal bike stowage (Right)

Colours, Materials and Finishes (CMF) and lighting

The inclusive design aspects of colours and patterns and the finishing of materials has received more focus in recent years (PAS 6463:2022). The inclusive design reviews highlighted the need to consider lighting and CMF, particularly for neurodiverse users.

The HF team prepared inclusive design requirements as an input to the CMF development. These were in addition to the minimum requirements for contrast between touchpoints and the background against which they are seen, between exterior doorways and the adjacent flooring, and passenger information required by NTSN PRM. For example, the specification of materials and finishes to consider visual noise and colours that support visual comfort (PAS 6463:2022).

The as-bid lighting concept in the vestibule, shown left in Figure 7, was thought by some users to result in excessive glare. For example, a deaf participant with an interpreter commented that it took more focus to understand and interpret questions during the inclusive design review and the high brightness gave her headaches. The main issue was thought to be the central panel light. After consideration and discussion with HS2 Ltd this has been replaced with spotlights, shown right in Figure 7, which results in a more appropriate brightness for partially-sighted users and deaf users who may lip read.



© Alstom SA 2025.

Figure 7: Vestibule lighting changes, as-bid on left and revised design shown right

Layout and Seating

The general layout of the saloon and the provision of seating, in particular seat pitch was very positively received by all participants e.g. in Figure 8. The seat pitch on HS2 is 870 mm compared to a typical 750 mm on UK trains. A partially sighted user, noted a “safe space, no hazards posed while walking from the main door area to the wheelchair seat”. Other users commented that the additional space around seats and ample storage supported users who may use assistance dogs or store equiptment.

.

Figure 8: Inclusive Design testing with assistance dogs

Conclusions

The Inclusive Design reviews helped define a methodology for the integration of inclusive design principles in the design of the new rolling stock, a baseline for future reviews, and potential changes to improve the overall inclusivity. Results were reviewed with system engineers, design engineers and the HS2 Ltd for feasibility of implementation. This also required the stakeholders to consider, for example, mandatory standards requirements and network operating restrictions. Implemented changes will be assessed at a later stage of the design process using the inclusive design review methodology to identify their impact on the overall inclusivity of the train. Implemented design changes have the potential to improve the travelling experience for all passengers taking into consideration those with mobility and sensory impairments, people who are neurodivergent and families.

References

Afacan, Y., & Erbug, C. (2009). An interdisciplinary heuristic evaluation method for universal building design. Applied Ergonomics, 40(4), 731-744.

Aya Bayramova, David J. Edwards, Chris Roberts, Iain Rillie, Enhanced safety in complex socio-technical systems via safety-in-cohesion, Safety Science, Volume 164, 2023.

BS16585-2:2017, Railway applications – Design for PRM use – General requirements.

BS700-6:2005, Design management systems – Part 6: Managing inclusive design – Guide.

Drew, C., The Double Diamond: 15 years on. Available from https://www.designcouncil.org.uk/news-opinion/double-diamond-15-years

Homepage - HS2 Learning Legacy.

PAS 6463:2022, design for the mind – Neurodiversity and the built environment – Guide.

Read, G J M, Naweed, A, Salmon, P M, Complexity on the rails: A systems-based approach to understanding safety management in rail transport, Reliability Engineering & System Safety, Volume 188, 2019, Pages 352-365.

Do rail workers still have a ‘feast-and-famine’ sleep pattern post-Covid?

Anna Vereker, Barbro Årnes & Claire Watt-Coombes

RSSB

SUMMARY

This paper presents key findings from the Rail Safety and Standards Board (RSSB) 2023 UK rail industry fatigue survey. A similar survey was run in 2018 and found that rail workers (especially shift workers) generally had a ‘feast-and-famine’ sleep pattern. We are now able to compare 2018 results with those from the 2023 survey to ascertain if this pattern still exists and to consider how features of fatigue risk management may have changed for rail workers in the post-covid period.

KEYWORDS

Fatigue risk management, Sleep, Shift work, Railway

Background and aim

In 2018 the RSSB facilitated a UK rail industry fatigue survey providing the first pan-industry insights into features of rail workers’ fatigue risk management. Key among these findings was that many rail workers (especially shift workers) reported a ‘feast-and-famine’ sleeping pattern which meant they were unlikely to be obtaining the quantity of sleep they needed (Basacik & Tailor, 2022). By 2023 rail industry companies were keen to repeat the survey to consider what elements of fatigue risk may have changed in the intervening five years, and particularly in the post-Covid period.

The 2023 fatigue survey provided an opportunity to ‘take the pulse’ of the industry’s fatigue risk management success from a rail workers’ perspective. This will help to inform the work of UK rail industry fatigue groups, the overall RSSB fatigue programme, and the wider Rail Health and Safety Strategy (RHSS) (RSSB, 2024).

Method

The original 2018 survey was reviewed and refined, with some further individual fatigue management questions added, and overall survey length reduced to under 20 minutes completion time. The questionnaire was digitised using the Microsoft Customer Voice™ application which was available on a variety of electronic devices and accessible via a web link and a quick-response [QR] code. Although a paper version of the questionnaire was available, no companies chose to use this. Participating companies from the previous survey were contacted. The survey was advertised via the industry fatigue groups and on RSSB’s social media platforms. In total 4143 survey responses were received, compared with over 7000 in 2018.

Results

Individual company analysis (N=24) and cross-industry analysis has been undertaken. This represented a variety of rail companies from passenger operations, freight operations, infrastructure and logistics, and supply companies.

Headline findings from the 2023 cross-industry analysis include:

- Around two in five rail workers said they had experienced a safety event at work where fatigue or alertness was a factor. This frequency looks to have increased from one in six in the 2018 survey and requires further analysis.
- 47% of respondents indicated that in the past month they have felt so tired at work that they have had to make some effort to stay awake once a week or more. In 2018 this frequency was 34%, and again requires further analysis.
- 72% of respondents admitted that they had driven a road vehicle while fatigued in the past 12 months. In 2018 this was 68%, and we need to investigate whether this is a statistically significant difference.
- 78% of respondents reported achieving less than six hours sleep between two night shifts, only 48% had less than six hours sleep between two day shifts, and 26% had less than six hours sleep between two days off. At first glance this suggests that the 'feast-and-famine' pattern of sleep continues for rail workers and requires further investigation.
- Shift working respondents provided a range of details about their most fatiguing shift in the past month including a self-rating on the Karolinska Sleepiness Scale (KSS), (Åkerstedt & Gillberg, 1990). 52% scored 7-9 (out of 9) on the scale, which is the level at which performance decrement has been shown in various task performance research (Kaida et al, 2006). The KSS is a central component of the Fatigue Risk Index (FRI) which is a fatigue prediction tool commonly used in the rail industry (QinetiQ & Simon Folkard, 2006).

Discussion

The initial cross-industry analysis seems consistent with results found in the 2018 survey, although further statistical analysis is ongoing. The fact that respondents continue to report lack of sleep between shifts (especially night shifts), driving a road vehicle when fatigued and working at potentially dangerous levels of fatigue suggests that rail fatigue risk management requires further development and enhancement. This highlights the importance of the RSSB's collaborative work as part of the Rail Health and Safety Strategy (RHSS), in assisting rail companies with their fatigue risk management activities.

Moreover, while the overall number of survey responses was lower in 2023 than in 2018, there was a broader spread of companies from different areas within the rail industry. This means the data (once fully analysed) can be used to help target, develop and implement fatigue risk management improvements.

References

Åkerstedt, T Gillberg, M. (1990) Subjective and objective sleepiness in the active individual. International Journal of Neuroscience, 52 (1-2).

Basacik, D & Tailor, A. (2022). A feast-and-famine pattern of sleep: Do railway staff get the sleep they need? Applied Ergonomics 102, 103711.

Kaida, K., Takahashi, M., Åkerstedt, T., Nakata, A., Otsuka, Y., Haratani, T., & Fukasawa, K. (2006) Validation of the Karolinska sleepiness scale against performance and EEG variables, Clinical Neurophysiology, 117(7).

RSSB (2024) The Rail Health and Safety Strategy (RHSS), Rail Health and Safety Strategy.

QinetiQ & Folkard, S. (2006) The development of a fatigue / risk index for shiftworkers: Research Report 446, HSE Books.

ECDP ETCS Pathfinder: The value of an In-Service Review

John Gunnell

Atkins Réalis

SUMMARY

The East Coast Digital Programme (ECDP) is the largest and most complex signalling upgrade programme in Network Rail history. It requires the deployment of wayside and onboard signalling and communication technologies across more than 100 miles of track and impacting 27 operators and multiple vehicle fleets with mixed classes, all of which must be fitted with European Train Control System (ETCS) onboard equipment to use the route. ECDP impacts on all elements of the Railway System.

London's Northern City Line (NCL), a provided the perfect project ECDP pathfinder to trial the system in operation on the live Network Railway with limited performance impacts on the East Coast Mainline.

It is important, and an industry requirement, to integrate Human Factors in major UK Rail projects with significant operational impact on various stakeholders. Effective Human Factors integration addresses the needs of technology development and business change projects within complex socio-technical systems.

The ECDP NCL 'pathfinder' commissioned in early 2024, introducing one of the only ETCS digital signalling systems in the UK. The early operational ETCS system service on ECDP has provided an opportunity to explore the system performance, and importantly, the users feedback on the changes to their ways of working.

KEYWORDS

Rail, signalling

In Service Review Methodology

The East Coast Digital Programme (ECDP) NCL Human Factors team followed a structured approach, summarised in Figure 1, to harness the know baseline understanding of the system to evaluate the actual operating conditions to aid the projects future priorities.

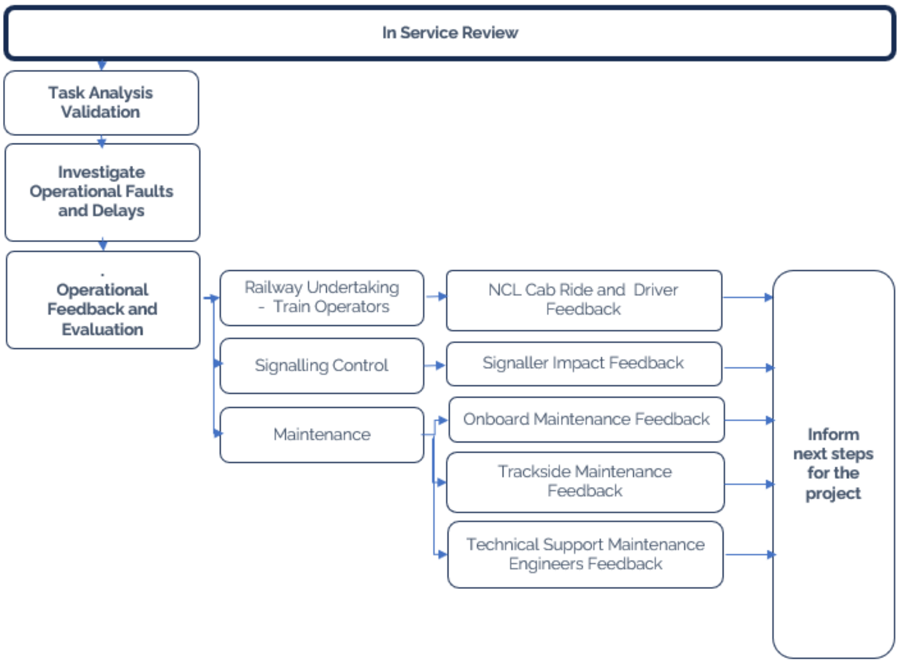


Figure 1: ECDP NCL In Service Review Approach

Work as defined vs work as done

The Task Analysis was developed based on operating principles defined by the project from recognised Industry Standards. With now the experience of training and operating the systems in a various condition, the Task Analysis could be validated with end users. A structured approach to review all normal, abnormal, and emergency scenarios provided a basis that not only the pathfinder scheme, but also for the broader system migration to come.

The clarified scenarios provided the actual human interface demands required with the new signalling system. A fundamental reference point for the ISR testing and analysis to build upon.

The differences, and shortcomings, of the documented processes compared to the reality, provided evidence for operational ambiguity. The validation exercise identified a series of operational demands and decision-making risks in an environmental context of pre-existing functions withstanding.

Reported Incidents and delay data

As expected with any implementation of such magnitude, the migration of new systems requires familiarisation and adaption during transition period. The uptake and effectiveness of both the user and the equipment tend to require learnings for reliability and effectiveness.

A set of ECDP ETCS-related reported incidents and operational delays were recorded by the client, Network Rail, on the pathfinder NCL. The main intention being to ensure the safety and performance of the network. By learning the root cause of the technical or human behaviours the project had dependable data to base improvements to people, process and on the pathfinder and future deployments.

As one would expect, the number of incidents and service delay minutes grew in correlation to the number of train Drivers trained and operating with the new systems.

Interestingly, even with the incident and root cause analysis taking place, the data reports continued to rise post completion of all Driver training on the NCL pathfinder. Certain pertinent issues were reoccurring time after time impacting the business confidence, and acceptance of system change.

The value of HF engineers integrating into investigations of in-service incident data reviews is to support understanding the ‘so what’ in specific ‘human error’ issues. Advising investigatory stakeholders that ‘human error’ is not the resulting issue, but the consequence, broadens the perspective of fault from simply the user’s actions to taking the time review the task interfaces in consideration. Understanding the complexities and risks in the operational pathfinder through user centred contextual data and feedback has challenged the design and integration of the system with new HF issues to address for future ECDP deployments.

Takeaways from the pathfinder In Service Review

The In-Service Review on the ECDP pathfinder project has highlighted the need for awareness and anticipation that the adoption of change will not be immediate. The value of focusing efforts on delving into the common teething issues is crucial to the long-term success of the bigger picture.

By harnessing the users of the early operating system to validate an understanding of the working interfaces it builds confidence in the benchmark system and best prepares the business readiness process, as the contextual environment has a large influence on the use.

Exploring the risk and extent of musculoskeletal disorders in UK heavy rail

Kirsten Huysamen, Owen McCulloch, Faye Bacon, Jemma Widdows, Barbro Arnes & Paul Leach

Rail Safety and Standards Board

SUMMARY

Musculoskeletal disorders (MSDs) are one of the leading causes of absence in the UK railway. The data required to address this issue is not readily available. The Rail Safety and Standards Board instigated a study to carry out the first industry-wide MSD survey to understand the prevalence and types of MSDs rail workers are experiencing and why this may be occurring. A railway-specific MSD survey was created using evidence from scientific literature and consultations with UK rail organisations, railway unions and the regulator, and the objectives of the survey. The survey underwent user testing to assess its readability, usability, and applicability to UK heavy rail. The survey was active for a duration of 5 months (July 2024 – December 2024) and open to all job roles contributing to heavy rail operations in the UK. The paper provides details on the survey development and initial findings.

KEYWORDS

Musculoskeletal disorders, musculoskeletal conditions, health and wellbeing, railway

Introduction

Musculoskeletal disorders (MSDs) are a significant issue in the UK rail industry, contributing to approximately 1 in 4 days of sickness absence and costing the industry approximately £89 million per year (ORR, 2019; RSSB, 2024). RSSB's Mental Wellbeing Survey (RSSB, 2021) found that nearly half of the rail workers that responded experienced backache (48%), upper body pain (54%) and lower limb pain (43%), in the past 12 months (4000 rail workers, 31 companies).

The UK rail industry currently has no reliable and consolidated data on the prevalence and types of MSDs, the affected job roles, and the tasks or risk factors contributing to their development. This makes it challenging to prevent and manage these disorders, develop targeted interventions, and determine where to focus effort.

The Rail Safety and Standards Board (RSSB) instigated a study to carry out the first industry-wide MSD survey to obtain the data required to understand the prevalence, type and risk factors of MSDs in UK heavy rail.

Method

Phase 1 produced a survey aligned to the project objectives, which are to:

- Understand the prevalence and types of MSDs rail workers are experiencing and the impact these have on workers' professional and personal lives.
- Identify the job roles affected by MSDs and, the tasks or risk factors contributing to the development of these disorders.

Phase 1 activities included a) review of scientific literature, standards, guidance and resources from reputable sources and b) consultations with UK rail organisations (rail operators, infrastructure and supply chain companies), railway unions and the regulator. A total of 11 consultations were conducted.

These activities gathered information on MSDs to inform the development of the survey, build an understanding of UK heavy rail job roles and MSDs within the industry, and identify validated or applicable MSD surveys. These findings, along with the project objectives, were used to create an online railway-specific MSD survey.

Phase 2 collected and analysed data from the survey and identified areas that need to be prioritised by the industry to improve musculoskeletal health. Before launching, the survey was user tested with 20 participants to evaluate its readability, usability and applicability to UK heavy rail. It was subsequently modified based on the feedback. This testing included participants with operational experience and utilised user scenarios to ensure a range of job roles were assessed.

The survey was active for five months (July 2024 to December 2024) and open to all job roles contributing to heavy rail operations in the UK. The data is currently in the process of being reviewed and analysed to identify key insights and trends, across all responses and within each individual job role. Several factors and relationships will be assessed, including but not limited to: a) the prevalence of trouble (aches, pain, discomfort) for each body part, b) the frequency of perceived high-risk tasks, c) exposure to MSD risk factors, d) likelihood of reporting, and e) perceptions of organisational culture related to MSDs. For each affected body part, aspects such as work as a causation factor, onset, sickness absence and medical interventions will be explored.

The remainder of this paper sets out findings from Phase 1, the survey development process, and initial survey data. As analysis is ongoing, data interpretation is not included.

Phase 1 Key Findings

Musculoskeletal disorders in UK heavy rail

Phase 1 findings revealed that MSDs are a significant issue in the UK heavy rail sector. Some rail organisations consulted attributed approximately 20% to 30% of their sickness absence to MSDs. Similar findings were reported by the Office of Rail and Road in 2019 (ORR, 2019). RSSB's Health and Wellbeing Data Hub, which collects health and wellbeing data from 21 companies across the GB rail industry, shows that MSDs and mental health issues are the two leading causes of sickness absence in the GB rail industry (McMahon et al., 2024; RSSB, 2024). The types of MSDs varied across organisations and railway sectors. Consultations most frequently highlighted lower limb and back issues, while RSSB's Mental Wellbeing Survey pointed to upper body concerns (RSSB, 2021). However, there was limited data for some job roles, the job roles most affected or the underlying causes.

The consultations confirmed that available detailed data on MSDs across the industry is limited, with some organisations having better insights than others. Organisations noted that this makes it difficult to identify trends and causation factors, leading to a reactive rather than proactive approach to MSD prevention and management.

Potential underreporting of MSDs in the rail industry was a key theme. Some organisations stated that employees tend to report MSDs only when they sustained an injury in the workplace requiring immediate medical attention or when pain becomes severe enough to prevent them from working. The latter makes it particularly challenging for organisations to identify causation factors and understand the prevalence and types of MSDs occurring. Moreover, there appears to be a discrepancy in reporting rates among job roles, with some reporting at a higher rate than others.

This means that some job roles may be overlooked, or their MSDs may go unnoticed. Organisations attributed underreporting to factors such as lack of MSD awareness, organisational culture (e.g., fear of job loss, the belief that aches, pain, and discomfort are part of the job, the perception that nothing will change) and gaps in reporting processes, for example, most organisation did not have a single, centralised process for reporting and capturing MSDs.

Contributing factors to musculoskeletal disorders

MSDs are considered multifactorial, often arising from a combination of factors rather than a single cause (Kiesel et al., 2024). There are numerous factors that can increase an individual's likelihood of developing an MSD (Punnett et al., 2004; HSE, 2002; da Costa et al., 2010; HSE, 2022, Kiesel et al., 2024). These can be grouped into three broad categories:

- Individual risk factors: Characteristics of an individual that can influence their susceptibility to developing an MSD (e.g., age, weight, health, fitness, lifestyle, past experiences etc.)
- Job & workplace risk factors: Physical demands and conditions of the job and workplace that contribute to the development of an MSD (e.g., manual handling, repetitive movements, awkward and static postures, bending and twisting, temperature, vibration etc.)
- Psychosocial risk factors: Cultural, social and psychological aspects of the job or working environment that are linked to the development of MSDs (e.g., job control, job demand, skill utilisation, recognition, breaks etc).

The consultations identified rail-specific MSD risk factors such as walking on ballast for extended periods, prolonged use of safety boots and working with damaged or old equipment. These were incorporated into the development of the survey.

Review of musculoskeletal disorder surveys

Numerous surveys were identified in the scientific literature and from reputable sources such as the Health and Safety Executive (HSE). Some of these surveys were designed for individuals with a preexisting MSD such as the Musculoskeletal Health Questionnaire (MSK-HQ) (Hill et al., 2016). Others were focused on identifying and understanding the potential musculoskeletal issues, if any, an individual has experienced in the past 3 or 12 months. These were typically based off the Nordic Musculoskeletal Questionnaire (NMQ) (Kuorinka et al., 1987). There were a few comprehensive MSD surveys, such as the Dutch Musculoskeletal Questionnaire (DMQ), which includes over 150 questions (Hildebrandt et al., 2001). These generally gathered information on potential musculoskeletal issues and exposure to MSD risk factors. There were also surveys that focused on a singular risk factor such as the HSE's Management Standards Indicator Tool (HSE-MS IT) which examines psychosocial risk factors in the workplace (Marcatto et al., 2014).

Developing a rail-specific musculoskeletal disorder survey

Using the Phase 1 evidence and project objectives, workshops were undertaken to identify the key information needed for the survey. Five categories of information were established, as shown in Table 1. None of the Phase 1 reviewed surveys fully captured this information, leading to the development of a heavy rail-specific MSD survey.

To develop the survey, each category of information was mapped to a relevant survey from Phase 1, where possible. The mapped surveys and their questions formed the foundation for the corresponding categories. Each question was then thoroughly reviewed and adapted to ensure relevance to UK heavy rail and alignment with the project's objectives. Non-relevant questions were excluded, and gaps were addressed by incorporating questions from other surveys or developing new ones. The main surveys used to develop the heavy rail-specific MSD survey, along with the mapped categories of information, are displayed in Table 1.

Table 1: The survey's five categories of information, with example data types and corresponding foundation surveys.

	Category	Type of information	Foundation surveys
1	General information	Demographic data, health, fitness, job role, experience, shift pattern, hours worked, contract type etc.	DMQ (Hildebrandt et al., 2001)
2	Musculoskeletal issues	Existing issues, the impact on work and daily life, characteristics of the issue (e.g., onset, frequency) etc.	Extended NMQ (Dawson et al., 2009) HSE's Body Mapping Tool (ORR, 2019)
3	High risk job tasks	Workplace tasks that could contribute to MSD development.	DMQ (Hildebrandt et al., 2001)
4	MSD risk factors	Exposure to job, workplace, and psychosocial MSD risk factors	DMQ (Hildebrandt et al., 2001) HSE-MS IT (Marcatto et al., 2014)
5	Organisation and reporting culture	Likelihood of reporting, MSD awareness, barriers to reporting, culture surrounding MSDs etc.	Not Applicable

The final version of the survey comprised six sections, as seen in Table 2. The final version was refined and modified based on feedback from rail organisations, railway unions and the user testing. The survey was also tested for technical feasibility at three rail organisations.

Table 2: The six sections of the UK heavy rail-specific MSD survey.

	Section title	Brief description of survey section
1	About you	Captures general information about the participant and their job.
2	Your musculoskeletal health	Identifies and explores musculoskeletal issues experienced in the past 12 months.
3	Your job tasks	Explores the tasks that may be causing aches, pain and discomfort at work.
4	Your job and work environment	Collects data on the exposure to factors linked to the development of MSDs.
5	Your work conditions	
6	Reporting in your organisation	Gathers information on the organisational culture surrounding MSDs and reporting thereof.

Phase 2 Initial Findings: Overall Survey Responses

This section presents some of the initial survey findings. As data analysis is still ongoing, this paper does not include data interpretation.

Demographic information

The survey received a total of 6,168 responses across various heavy rail job roles, with some roles being more represented than others. As seen in the Figure 1, job roles working on a train (e.g., train driver, conductor), had the highest response rate, whereas job roles working in a control or signalling location (e.g., signaller, controller) had the lowest response rate.

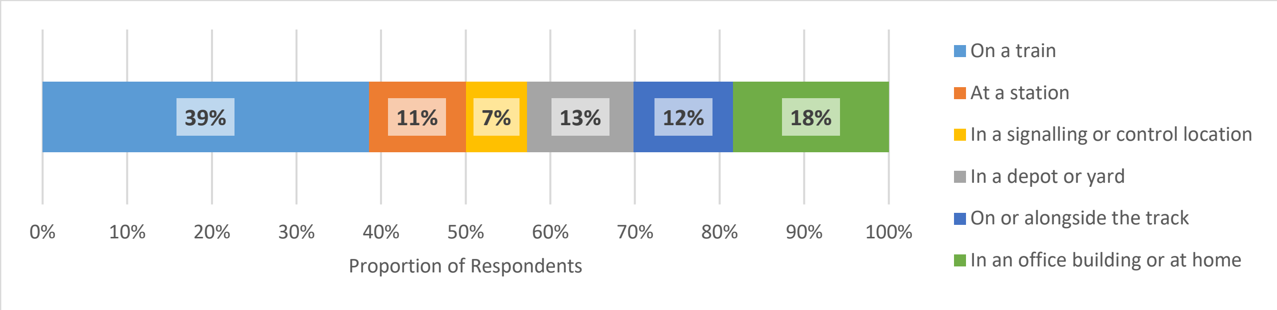


Figure 1: Proportion of MSD survey respondents by work location.

The age, experience level, and weekly working hours of respondents are presented in the Table 3. Most respondents were shift workers (77%), and on permanent employment contracts (93%). The majority were male (75%), while 24% were female and 1% indicated ‘Other’ or ‘Prefer not to say’. Most respondents reported being in good physical and mental health and physically fit.

Table 3: Proportion of MSD survey respondents by age, experience level and hours worked.

Age		**Rail industry experience**		**Job role experience**		**Hours worked per week**	
Years	**% Responses**	**Years**	**% Responses**	**Years**	**% Responses**	**Hours**	**% Responses**
< 20	0.4%	< 1	4%	< 1	7%	0-20	5%
21-30	8%	1-5	18%	1-5	30%	21-30	3%
31-40	19%	6-10	19%	6-10	20%	31-40	56%
41-50	26%	11-15	11%	11-15	11%	41-50	30%
51-60	36%	16-20	14%	16-20	12%	51-60	5%
> 61	10%	> 21	34%	> 21	21%	> 61	1%

* *% Responses refer to the proportion of respondents*

Musculoskeletal issues

Majority of respondents reported having trouble (aches, pain, discomfort) with one or more of their body parts in the past 12 months (94%). The most affected area was the lower back (64%), followed by the knees (48%), shoulder (46%) and neck (42%). The least reported were upper arms, elbows, and lower legs/calves. Further details are provided in Figure 2. The lower back was the most reported body part to cause trouble across all job roles with over 20 responses.

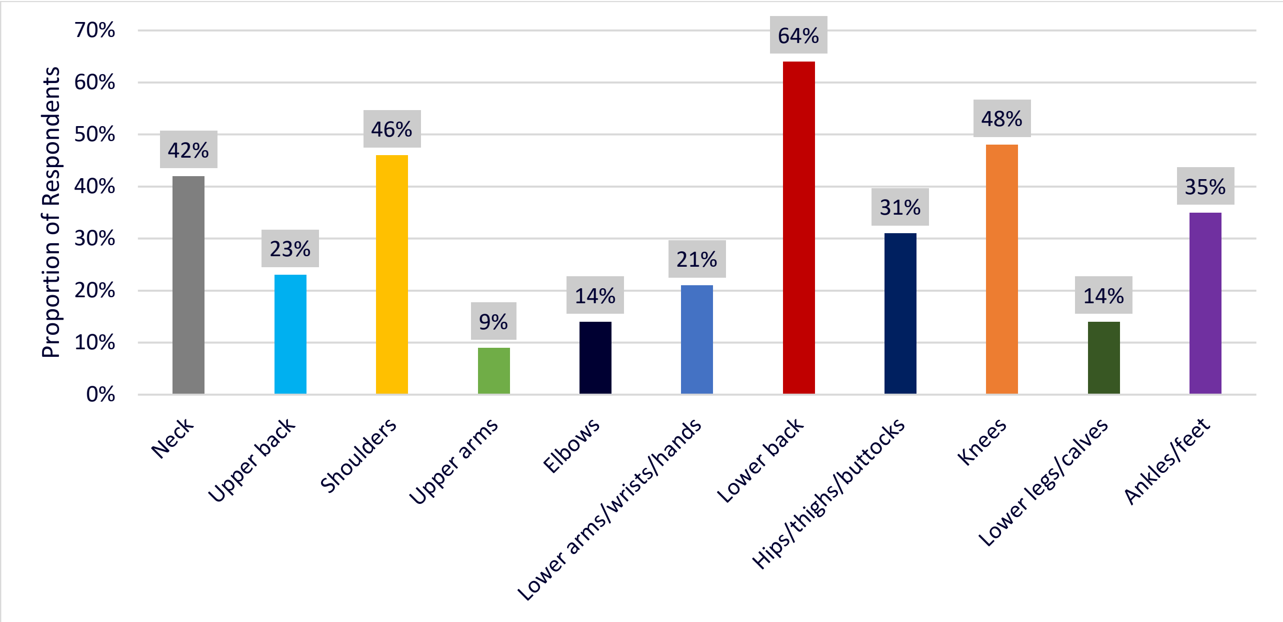


Figure 2: Proportion of MSD survey respondents reporting trouble (aches, pain, discomfort) by body part in the past 12 months.

Reporting and organisational culture

The findings indicate that musculoskeletal issues impacting an individual's ability to perform their duties may be underreported. 12% of survey respondents who indicated experiencing a musculoskeletal issue that affected their work duties in the past 12 months selected "no" when asked if they had reported it to their company. This trend was consistent amongst job roles, though the potential underreporting rate varied, with some job roles showing estimates of 20-25%. For example, for trackside 'Maintenance Workers', 57% indicated experiencing a musculoskeletal issue affecting their work, yet only 32% reported such issues to their company.

The top 5 reported reasons for not reporting were:

- Worrying about management responding negatively (38%)
- Believing that aches, pain and discomfort are part of their job (35%)
- Believing nothing will be done if they report a musculoskeletal issue (31%)
- Preference to not change their duties or job (28%)
- Being concerned about potential job loss (28%)

The primary reasons for not reporting varied slightly amongst job roles. For instance, among mainline 'Train Drivers', the most common reason was 'Worrying about management responding negatively' (50%), followed by 'Believing nothing will be done if they report a musculoskeletal issue' (46%). In contrast, for trackside 'Maintenance Workers', the top reason was 'Believing that aches, pain and discomfort are part of the job' (51%), with 'Worrying about management responding negatively' (44%) as the second most cited reason.

Phase 2 Initial Findings: Job Role Specific Responses

This section details initial findings for three job roles: mainline 'Train Drivers' (1083 responses), trackside 'Maintenance Workers' (390 responses) and lever frame box 'Signallers' (137 responses).

Musculoskeletal issues

Respondents across all three job roles most frequently reported experiencing trouble (aches, pain, discomfort) with their lower back in the past 12 months. The second and third most reported areas differed across the job roles, as seen in Table 4.

Table 4: The top 3 body parts most frequently reported to cause trouble for mainline 'Train Drivers', trackside 'Maintenance Workers' and lever frame box 'Signallers'.

	Most Frequently Reported		**Second Most Reported**		**Third Most Reported**	
Job role	**Body part**	**% Responses**	**Body part**	**% Responses**	**Body part**	**% Responses**
Mainline 'Train Drivers'	Lower back	71%	Shoulders and Neck			53%
Trackside 'Maintenance Workers'	Lower back	67%	Knees	56%	Shoulders	42%
Lever frame box 'Signallers'	Lower back	58%	Neck	45%	Shoulders	42%

** For mainline 'Train Drivers', the neck and shoulders were reported equally.*
** % Responses refer to the proportion of respondents.*

Lower back findings across job roles

More than half of mainline 'Train Drivers' (58%) and trackside 'Maintenance Workers' (61%) attributed their lower back trouble to work. This was slightly lower for lever frame box 'Signallers', with 30% reporting work as the cause. For all three job roles, at least 30% of respondents were unsure whether work caused their trouble. This suggests that a small proportion of respondents were confident that work was not a contributing factor (12%, 9% and 27% for mainline 'Train Drivers', trackside 'Maintenance Workers', and lever frame box 'Signallers', respectively).

A slightly higher proportion of trackside 'Maintenance Workers' (33%) took sickness absence because of their lower back trouble compared to mainline 'Train Drivers' (29%), while lever frame box 'Signallers' (19%) had the lowest rate. Lower back trouble appears to have a greater impact on the work duties of trackside 'Maintenance Workers', with 50% reporting it affected their work duties, compared to 34% for mainline 'Train Drivers' and 25% for lever frame box 'Signallers'.

More than half of respondents reported taking medication to alleviate pain or discomfort associated with their lower back trouble, while over 49% sought help from a medical professional (mainline 'Train Drivers': 60%, trackside 'Maintenance Workers': 54%, lever frame box 'Signallers': 49%). Most respondents stated their trouble was aggravated by work, was a recurring issue with a gradual onset, and that it impacted their leisure activities.

Perceived high risk tasks

Survey respondents were asked to describe up to two tasks that are likely to cause aches, pain, or discomfort at work, if applicable. The most frequently reported tasks included:

- Mainline 'Train Drivers': Sitting for long periods, operating the DSD pedal, driving the train and, boarding and alighting the train.
- Trackside 'Maintenance Workers': Manual handling, specifically lifting and carrying, and walking on or alongside the track on ballast.
- Lever frame box 'Signallers': Operating the signal and points levers, specifically pulling and pushing the levers.

MSD risk factors

Participants were asked to report the frequency of their exposure to various job, workplace, and psychosocial risk factors. Table 5 highlights the most frequently encountered risk factors across all three job roles. The top 5 job risk factors and top 3 psychosocial risk factors are presented, while workplace risk factors are included if more than 50% of respondents reported frequent exposure.

Table 5: The top 5 job risk factors, top 3 psychosocial risk factors, and workplace risk factors (reported by more than 50% of respondents as frequently encountered) for mainline 'Train Drivers', trackside 'Maintenance Workers', and lever frame box 'Signallers'.

Mainline 'Train Drivers'		**Trackside 'Maintenance Workers'**		**Lever frame box 'Signallers'**	
Job risk factors					
Sitting for long periods	98%	Prolonged walking or working on irregular, uneven or slippery surfaces	86%	Doing repetitive movements with the arms, hands or fingers.	89%
Performing repetitive movements with the arms, hands or fingers	89%	Frequent bending or twisting of the back/trunk	71%	Frequent bending or twisting of the back/trunk	66%
Repeating the same movements for long periods	80%	Standing for long periods	69%	Exerting great force with the arms, hands or fingers	61%
Performing repetitive movements with the legs or feet	80%	Doing repetitive movements with the legs or feet	60%	Repeating the same movements for long periods	61%
Frequent bending or twisting of the wrist or elbow	63%	Handling heavy loads (5kg - 20 kg)	56%	Handling heavy loads (5kg - 20 kg)	60%
Workplace risk factors					
Rarely experience tools or equipment that make work easier or more comfortable	64%	Working outdoors	93%	Rarely experience tools or equipment that make work easier or more comfortable	77%
Working with damaged, worn or old equipment, plant or locomotives	53%	Working in very cold or very hot temperatures	69%		
		A workstation, working environment or tools that cannot be adjusted	57%		
Psychosocial risk factors					
Concentrating very hard and paying close attention	98%	Concentrating very hard and paying close attention	75%	Performing the same task/ actions over and over again	92%
Performing the same task/ actions over and over again	95%	Performing the same task/ actions over and over again	58%	Concentrating very hard and paying close attention	92%

Staff aren't always consulted about a change at work	61%	Staff aren't always consulted about a change at work	50%	Limited opportunities for social interaction at work	82%

** % refers to the proportion of respondents who frequently encountered the risk factor.*

Conclusion and next steps

This study highlights the prevalence of musculoskeletal health issues in UK heavy rail, with lower back trouble identified as the most affected body area. These issues not only contribute to sickness absence but also negatively affect worker performance. Findings suggest that work-related factors play a role in both the development and aggravation of these conditions. MSDs are likely underreported, which may be influenced by organisational and reporting culture factors. This underreporting makes it challenging to identify trends and causation factors.

Although several MSD surveys exist, none fully captured the information required for this project, which led to the development of a heavy rail-specific MSD survey. It is likely that other safety-critical industries will also need to develop their own industry-specific surveys to capture relevant data. This paper outlines a method and process to create such a survey, which can be applied by other industries conducting similar studies.

The next steps include completing the data analysis and publishing the findings. The results will provide the industry with valuable insights, helping organisations understand the severity and extent of MSDs in the workplace. These findings will also highlight priority areas that need to be addressed to improve the musculoskeletal health of the UK heavy rail workforce. Additionally, the findings will inform the development of an industry-wide MSD strategy aimed at reducing the prevalence and impact of MSDs across various job roles.

References

da Costa, B. & Vieira, E. (2010). Risk factors for work-related musculoskeletal disorders: A systematic review of recent longitudinal studies. American Journal of Industrial Medicine, 53(3): 285-323.

Dawson, A., Steele, E., Hodges, P. & Stewart, S. 2009. Development and test-retest reliability of an extended version of the Nordic Musculoskeletal Questionnaire (NMQ-E): a screening instrument for musculoskeletal pain. The Journal of Pain, 10(5): 517-526.

Hildebrandt, V., Bongers, P., van Dijk, F., Kemper, H. & Dul, J. (2001). Dutch Musculoskeletal Questionnaire: description and basic qualities. Ergonomics, 44(12): 1038-1055.

Hill, J., Kang, S., Benedetto, E., Myers, H., Blackburn, S., Smith, S., Dunn, K., Hay, E., Rees, J., Beard, D., Glyn-Jones, S., Barker, K., Churchman, D., Ellis, B., Fitzpatrick, R., & Price, A. (2016). Development and initial cohort validation of the Arthritis Research UK Musculoskeletal Health Questionnaire (MSK-HQ) for use across musculoskeletal care pathways. BMJ Open, 6:e012331. doi:10.1136/bmjopen-2016-012331.

HSE. (2002). Upper limb disorders in the workplace. United Kingdom: Health and Safety Executive.

HSE. (2022). Work-related musculoskeletal disorders statistics in Great Britain, 2022. United Kingdom: Health and Safety Executive.

Kiesel, K., Matsel, K., Bullock, G., Arnold, T. & Plisky, P. (2024). Risk Factors for Musculoskeletal Health: A Review of the Literature and Clinical Application. International Journal of Sport Physical Therapy, 19(10): 1255-1262.

Kuorinka, I., Jonsson, B., Kilbom, A., Vinterberg, H., Biering-Sørensen, F., Andersson, G., & Jørgensen, K. (1987). Standardised Nordic questionnaires for the analysis of musculoskeletal symptoms. Applied Ergonomics, 18(3), 233–237.

Marcatto, F., Colautti, F., Filon, F., Luis, O. & Ferrante, D. (2014). The HSE Management Standards Indicator Tool: concurrent and construct validity. Occupational Medicine, 64(5): 365-371.
McMahon, N., Regan, L., Masetti, I, & Sobun, N. (2024). The Rail Industry Health and Wellbeing Performance Indicators Report: Quarter 1, 2024–25. London: Rail Safety and standards Board.
ORR. (2019). Closing the gap on health: ORR's review of health risk management in the rail industry for 2014-2019. United Kingdom: Office of Road and Rail.
Punnett, L. & Wegman, D. (2004). Work-related musculoskeletal disorders: the epidemiologic evidence and the debate. Journal of Electromyography and Kinesiology, 14(1): 13-23.
RSSB. (2021). The Rail Industry Mental Wellbeing Survey. London: Rail Safety and Standards Board.
RSSB. (2024). Annual Health and Safety Report 2023/24: A snapshot of GB rail's health and safety performance. London: Rail Safety and Standards Board.
RSSB. (2024). Annual Health and Safety Report 2023/24: A snapshot of GB rail's health and safety performance - Health and Wellbeing Data Appendix. London: Rail Safety and Standards Board.

Improving users’ performance and safety at Station Pedestrian Crossings: a qualitative study

Yanna Carli[1], Christopher Paglia[1]

[1]IRT Railenium

SUMMARY

In order to improve safety at Station Pedestrian Crossings (SPC), a project was launched to develop a new safety equipment using a behavioural approach. This paper presents a qualitative study using focus groups to identify promising safety equipment concepts based on the theories of situation awareness and perception/action loop. Thirteen participants took part in activities and discussions involving scenarios simulation, mock-ups and questionnaires. Safety equipment elements embodying the theoretical aspects studied were analysed and their design improved to identify the most promising ones regarding pedestrians’ performance and safety, i.e. the ones inducing safe behaviours. Combinations of the latter were then evaluated to identity the best safety equipment concepts that will serve for the development of refined prototypes, the efficacy of which regarding safety at SPC crossings will be evaluated on a mixed-reality testing platform developed on purpose.

KEYWORDS

station pedestrian crossing, railway safety, safety equipment, focus group, situation awareness

Introduction

Station Pedestrian Crossings (SPC) enable pedestrians to cross rail tracks at grade in stations where there are no footbridges or underpasses. These crossings involve walking directly on rail tracks, which exposes users to a risk of collision with passing trains. In 2023, a project was launched to improve pedestrians’ safety at SPC. Major risks reported at SPC were the ‘another train coming’ risk, distraction of SPC users and social influence, which prompted the development of an approach based on users’ behaviour understanding. A first objective of the project was to investigate human behaviours and cognitive processes involved in the action of crossing a SPC and which could serve as levers for safety improvement. A literature review helped to identify situation awareness (SA) (Endsley, 1995) and the perception/action loop (Gibson, 1979) as relevant factors to study. A second objective was to develop a new safety equipment, the design of which would be guided by the theoretical factors identified.

Method

A qualitative study using focus groups appeared to be a suitable method to analyse human behaviours and to get users involved in the design process of safety equipment. Regarding the design of safety equipment concepts, the use of simulation and mock-ups appeared to be a good approach to make participants interact with design elements. Indeed, mock-ups allow for a simplified representation of environments and situations, which can be as effective as a detailed working prototype and less costly (Bell, 2007). Furthermore, it avoids the risks inherent to experimentations in real-life conditions (Watkins et al., 2008). Additionally, Karlsson (1996) indicates that the use of mock-ups and scenarios helps users express their needs. Four focus groups were therefore organised to generate the safety equipment concepts. Activities involving scenarios simulation and mock-ups for design improvement were developed. The focus groups sessions took place in a meeting room, around a table, and were led by a human factors expert (designated as *moderator* in the following) supported by an assistant for note-taking and material distribution.

Participants

The participants were thirteen unpaid volunteers (average age = 38,5±14,3, seven women, eight men), speaking French fluently, recruited through word of mouth and leafleting at train stations. They were users of French train stations, and non-experts of the railway sector and HOF domain. Twelve of them were familiar with pedestrian level crossings. They all had a good hearing and vision. The same panel was reconducted for each focus group to be efficient from one session to another and to go further in the analyses.

Materials

A 3D-reduced model of a small train station equipped with two platforms, a portion of a double-track railway, a SPC and various equipment (shelter, digital departures/arrivals board, dustbin, bench, ticket vending machine, platform numbers), as well as active safety device elements, were used to simulate scenarios (Figure 1). Furthermore, 2D mock-ups of a SPC and safety equipment elements were utilised by participants to generate design propositions (Figure 2). The 2D mock-ups of safety equipment elements were a mix of existing and innovative solutions designed on purpose by a partner on the project.

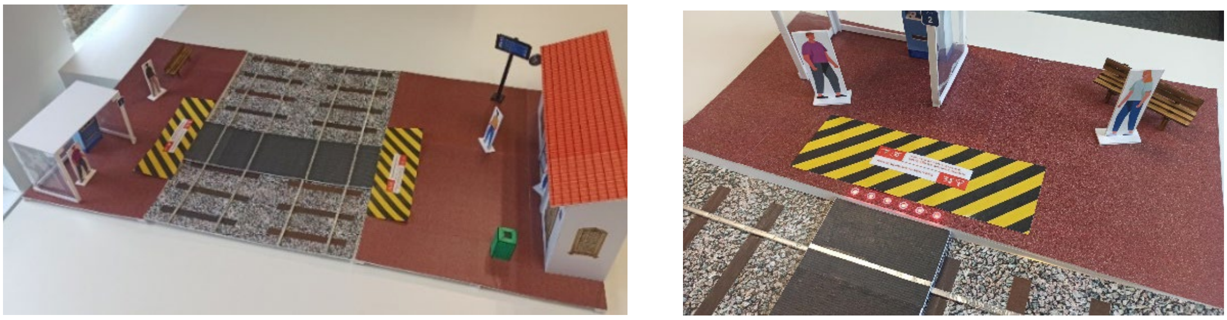


Figure 1: 3D-reduced model of a train station composed of two platforms, a portion of a double-track railway, a SPC (left); model of a safety device element shown activated (right).

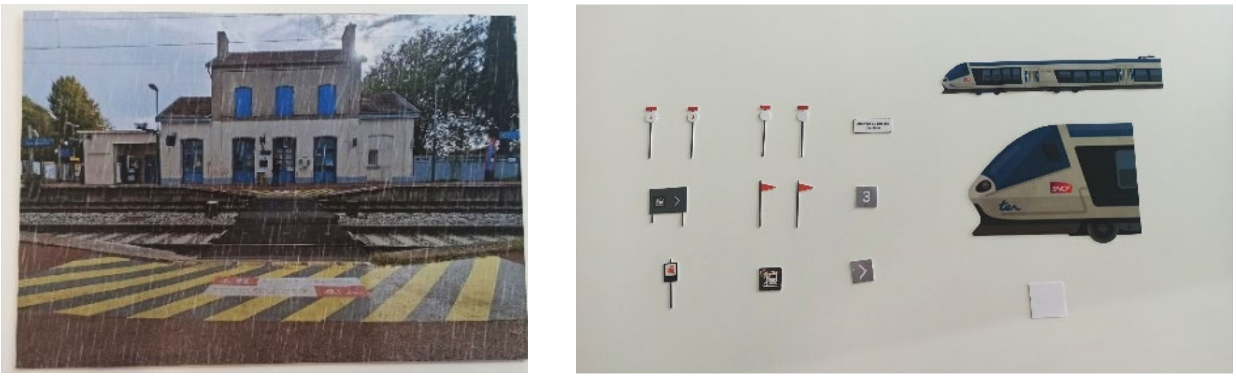


Figure 2: Example of the 2D mock-ups of a SPC (left) and safety equipment elements (right) used in the design improvement activity.

Procedure

The first three focus groups aimed at studying separately each theoretical aspect (levels of SA, perception/action loop) and identifying promising safety equipment elements enhancing these cognitive processes. The purpose of the fourth focus group was to compare combinations of equipment elements derived from the results of the previous focus groups, rank them and identify the most promising concepts regarding how well they guide users' behaviour and guarantee their safety at SPC. Each focus group was divided into two sessions, each session receiving half of the panel of participants. The groups of participants in each session varied from one focus group to another, which prevented the same participants taking the lead, supported an up-to-date understanding of the activities objectives, and guaranteed a renewal of ideas and results. The focus groups organisation is described in Figure 3.

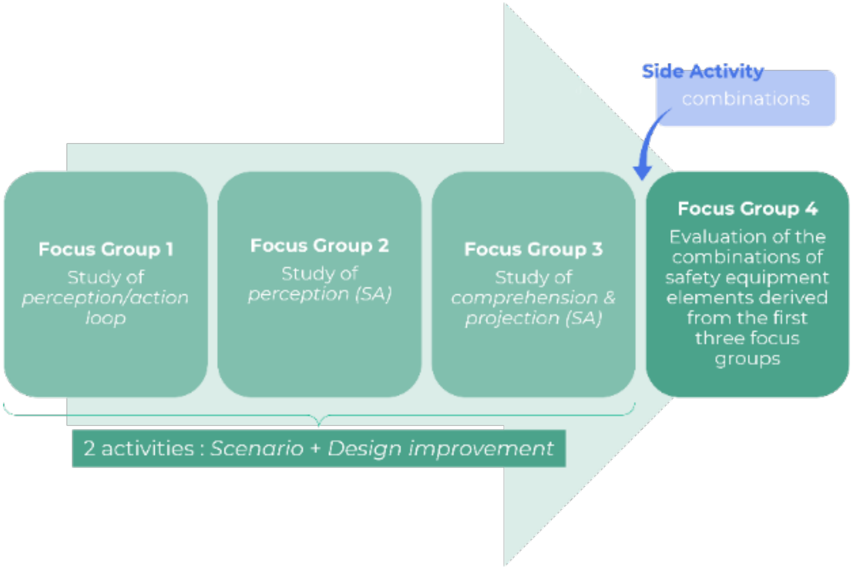


Figure 3: Focus groups organisation

The first three focus groups followed the same procedure. The idea was to present various safety equipment elements to the participants through scenarios, analyse the behaviours these elements induced and improve their design. In each focus group, the safety equipment elements used and the behaviours analysed were adapted in line with the theoretical aspect studied. The first focus group aimed at studying the *perception/action loop* (Gibson, 1979), i.e. see if safety elements would make participants stop spontaneously before the SPC when perceiving them. The second focus group studied the first level of SA (Endsley, 1995), *perception,* i.e. see if participants would detect well safety elements and interpret accurately the information they deliver (i.e. crossing ban). The third focus group focused on analysing the second and third level of SA, *comprehension and projection*, i.e. see if safety elements would help the participants understand different traffic situations at SPC (stopping train, passing train, 'another train coming' situation) and project the risk inherent to it (risk of collision with a train). The participants were not informed of the purpose of the study and did not know what activities to expect before the first focus group. The sessions were audio recorded. In a first activity (Activity 1), the participants were presented a **scenario** in which a fictional character had to achieve a goal inside a 3D-modelled train station environment implying the crossing of a SPC equipped with a safety equipment element shown activated (i.e. prohibiting the crossing of the SPC and announcing the approach of a train) (Figure 2, right). An example scenario was: "*Luke enquires for his train departure platform by looking at the departures board. The train is announced on platform 2. Luke is going on the corresponding platform to wait for his train.*". The participants had to explain what behaviours and actions the fictional character should put in place to achieve the goal, first through individual writings and then through a group discussion. This followed a *nominal group* approach (Gustafson et al., 1986) which gives everyone the opportunity to give their opinion and prevent the participants from getting biased by the others' answers. The group discussions would help the participants identify potential design flaws in the equipment element presented, related to the scenario situation, and give them ideas for the next step.

In a second activity (Activity 2), the participants were divided into groups of two or three people and given 2D-mock-ups of a SPC and safety equipment elements (Figure 2) to propose a **design improvement** of the safety equipment element encountered in the first activity. The technical functioning of the elements was presented on a screen in animated versions and left available for consultation. The design improvement was compelled by several instructions: the proposition should trigger the adequate behaviours / actions (relative to the theoretical aspect studied), a maximum of two safety equipment elements should be used, the participants had the possibility to suggest only one new element (drawn on a blank card), environmental and/or situational factors should be taken in account (weather conditions, trains traffic). The design proposition of each group was evaluated by the other participants through individual questionnaires and then discussed orally. The analyses of the questionnaires and discussions served to identify the most promising design propositions, and therefore, the most promising safety equipment elements for each theoretical aspect studied.

Between the third and fourth focus groups, a side activity was carried out by the moderator and other human factors experts on the project. The purpose of the activity was to give a more generic form to the safety equipment elements composing the best design propositions identified earlier in each focus group. This ensured that theoretical principles inherent to the elements were validated ahead of the fourth focus group. Combinations of these elements were generated based on the results of a subsidiary quantitative study conducted in parallel by partners on the project. These results indicated that most efficient safety equipment concepts should integrate elements of *perception/action* and *perception* (SA) (Principle 1) or elements of *perception/action*, *perception* (SA) and *projection* (SA) (Principle 2). From all the combinations identified, only the relevant ones were kept (i.e. redundancies and incoherent matches were removed).

In the fourth focus group, the combinations of generic safety elements defined in the side activity were shown in animated versions to the participants and in random order. Some combinations were shown with different design examples. Each combination was associated with a letter. Participants were invited to rank individually the combinations in relation to their potential to enhance performance and safety. For that purpose, each participant was given a ranking grid with repositionable letters corresponding to the combinations. Ranking was adjusted by the participants after each combination was shown. The four best ranked concepts were kept for the subsequent phases of the project.

Measures and analyses

The participants' individual writings from Activity 1 were analysed using task analysis and content analysis with the aim to assess their intuitive behaviour relative to the theoretical aspects embodied by the safety equipment element activated on the train station 3D model. The task analysis helped defining the expected behaviours and actions to be observed when crossing a SPC. The occurrence of specific combinations of verbal and thematic components in the participants' writings would indicate that they adopted the expected behaviours in front of the equipment element presented on the train station model. This gave an indication of the efficacy of the latter. For example, in the focus group dealing with *perception*, if the behaviours and actions expected corresponded to the SPC crossing task steps "Look at the signage" and "If signage is active, do not cross the SPC", a combination of verbal expressions equivalent to "Luke looks at the flashing light" and "He waits for the flashing light to stop before crossing" was expected in the writings. The content analysis of the group discussions allowed for a better understanding of the behaviours observed in the individual step and gave a first insight into good and bad design considerations.

The questionnaire filled in in Activity 2 used 6-points and 3-points Likert scales, as well as open questions. Several aspects of the design propositions produced were investigated: the efficacy of the solution (Dean et al., 2006 ; Cheeley et al., 2018) (i.e. in the present case, to what extent the proposition allows for a safe crossing), the theoretical aspect involved (*perception/action loop*: detectability and affordance of the solution; *perception*: detectability and interpretability of the solution; *comprehension and projection*: detectability and interpretability of the solution, comprehension of the global situation and projection of its evolution), the perceived usefulness of the solution (i.e. "the degree to which a person believes that using a particular system would enhance his or her job performance", Davis, 1989) and its perceived ease of use (i.e. "the degree to which a person believes that using a particular system would be free of effort", Davis, 1989). The total number of questions differed from one focus group to another, depending on the theoretical aspect studied. Each question was assigned a score contribution. These were summed to calculate the questionnaire total score. Scores of all questionnaires related to the same design proposition were averaged. The comparison of the average scores of each design proposition determined what the best ranked proposition in each focus group was. If the average scores of two propositions were equal or too close, a review of the comments made by the participants on the solutions (in the questionnaire open questions and during the group discussion) helped to decide the best ranked. It was made sure that these complied with the instructions given.

At the end of the fourth focus group, the participants' individual rankings of the combinations of generic safety equipment elements were averaged to obtain the four most relevant safety equipment concepts.

Limitations

- The panel of participants being the same for each focus groups, the participants might have expected the activities planned if they were similar to those of the previous focus group(s).
- The participants were not in real-life conditions and could only rely on the 3D model which has limits in terms of realism. Their projection in the scenarios given was therefore limited and different from what it would have been in real-life conditions, and so were their oral and written answers.
- There is a possibility that participants might not have fully understood the design propositions of other groups, which has an influence on the open answers and scores given.
- Participants have different behaviours regarding scores allocation which questions the use of average scores to obtain a ranking of the design propositions. This was balanced by the consideration of participants' written and oral comments.

Results

Seven safety equipment elements were extracted from the most promising design propositions identified in the first three focus groups. The side activity allowed to derive six generic safety elements from the elements previously extracted, two per studied thematic. Among all the combinations generated from these generic elements, thirteen were kept and were then ranked by the participants in the fourth focus group. The combination rankings of all the participants were found to be similar, with four safety equipment concepts standing out by far from the average ranking and bringing a quasi-consensus in the participants' discussions that followed regarding their potential to enhance safety and improve users' behaviour at SPC. The multiple results and the discussions pinpointed several design guidelines for the development of a safety equipment at SPC.

Conclusion

The methodology used helped to generate a wide range of solutions with the potential to improve safety, from single safety elements addressing target individual human factors such as attention, perception and comprehension, to combined equipment enhancing users' overall situation awareness. The study of each theoretical aspect individually helped to build technological assets.

The accumulation of these assets through the succession of focus groups ensured an improved user's performance and safety at SPC. Users' involvement allowed for the expression of various viewpoints, going beyond a technical expertise of the studied systems. Four generic safety equipment concepts were identified, with a quasi-consensus between participants regarding their potential to enhance safety and induce adapted users' behaviours. The concepts were not opposite to each other but were rather an improved version of the concept of lower rank. The next step will be to combine the results of the qualitative study with those of quantitative and technical studies to settle on final improved concepts to be tested in the next phases of the project, that are tests in a mixed-reality environment. In order to implement mixed reality, a testing platform has been developed. It consists of a virtual environment synchronised with a scale 1:1 modular physical platform (Figure 4). The testing platform is set to be adaptative so that any environment can be modelled in detail. For the purpose of the project, part of a train station is being recreated in both virtual and physical environments including two trains station platforms, a portion of a double-track railway and a SPC. Behavioural, physiological and kinematic data will be collected from the participants and recorded using a VR headset and cameras. These measures will help evaluate more precisely the impact of the safety equipment concepts on users' behaviour and safety at SPC, before an implementation in real-life settings.

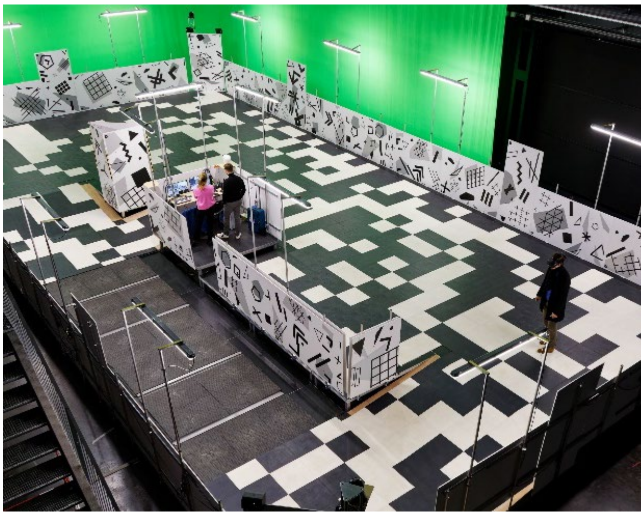


Figure 4: Mixed-reality testing platform developed

References

Bell, K. (2007, May). Mock-ups: Giving hospital clients the ultimate reality check. *Healthcare Design*.

Cheeley, A., Weaver, M. B., Bennetts, C., Caldwell, B. W., & Green, M. G. (2018, August 26). A Proposed Quality Metric for Ideation Effectiveness. *30th International Conference on Design Theory and Methodology, 7.*

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 319.

Dean, D. L., Hender, J. M., Rodgers, T. L., & Santanen, E. (2006). "Identifying good ideas: constructs and scales for idea evaluation", *Journal of Association for Information Systems*, *7*(10), 646-699.

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32-64.

Gibson, J. J. (1979). *The ecological approach to visual perception: classic edition*. Psychology Press.

Gustafson, D. H., Delbecq, A. L., & Van de Ven, A. H. (1986). *Group techniques for program planning-a guide to nominal group and Delphi processes.*
Jones, D. G., & Endsley, M. R. (2004). Use of Real-Time Probes for Measuring Situation Awareness. *The International Journal of Aviation Psychology*, *14*(4), 343-367.
Karlsson, M. (1996). *User Requirements Elicitation-A Framework for the Study of the Relation between User and Artefact*. Chalmers University of Technology.
Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science, 46*(2), 186-204.
Watkins, N., Myers, D., & Villasante, R. (2008). Mock-Ups as "Interactive Laboratories": Mixed Methods Research Using Inpatient Unit Room Mock-Ups. *HERD: Health Environments Research & Design Journal*, *2*(1), 66-81.

Integration of Human and Organisational Factors in Railway Systems Lifecycle Processes

Nora Balfe[1], Virginie Papillault[2] & Anna Windischer-Unterkircher[3]

[1]Iarnród Éireann Irish Rail, [2]International Union of Railways (UIC), [3]Swiss Federal Railways (SBB)

SUMMARY

Integration of human factors in railway change management and systems engineering is an increasingly important topic, not least because of emerging regulatory requirements in the area. This paper will describe one model developed collaboratively in the International Union of Railways (UIC) Human and Organisational Factors Working Group (HOFWG) which maps 12 human and organisational factors steps to the 12 steps in the V-cycle model described in the railway Reliability, Availability, Maintainability and Safety (RAMS) standard (EN50126). The aim is to provide a framework, aligned with existing engineering approaches, to communicate how human and organisational factors can be integrated into a railway change project.

KEYWORDS

Human factors integration, rail human factors, V-cycle

Introduction

Consideration of human and organisational factors (HOF)[1] during design and development of new railway systems is essential for ensuring the safe and efficient operation of the system. Generally speaking, HOF are not yet sufficiently addressed and are often applied implicitly, sometimes insufficiently considered and can be limited to physical ergonomics. However, taking HOF into account offers advantages in terms of both performance and safety. MANPRINT is the original model for human factors integration (HFI) but it doesn't fit well with modern railway projects (Balfe, 2023), and so does not support understanding of the human factors inputs, approaches and outputs among key stakeholder groups. For example, some elements such as 'Personnel' are usually handled by human resources departments while the core area of Human Factors Engineering (HFE) is underdefined. More recently, the European Union Agency for Rail (ERA) has published guidance on integrating HOF in change management (ERA, 2023). This guidance draws from the 'Human Factors Case' methodology used by Eurocontrol (2011) and the human-centred design process described in ISO9241-210 (2019). The work described in this paper is complementary to the ERA guidance, and intended to describe the same type of approach, but in the context of the widely used railway system V-cycle model as described in EN50126-1 (2017).

The work was undertaken by the UIC HOFWG in conjunction with the Performance, reliability, availability, maintainability and safety (PRAMS) HOF subgroup of Europe's Rail Joint Undertaking (ERJU) and resulted in the production of a document titled 'Integration of Human and Organisational Factors (HOF) in railway system design. In many European countries, there are regulatory requirements that mandate the application of EN 50126 in the development of railway applications. The norm underlines that the inclusion of human factors in the phases of the

[1] Note that the term 'human and organisational factors' is increasingly used in European railways where it is defined in the same way as 'human factors'. The terms are used synonymously in this paper.

development process is essential. However, it does not describe how human factors integration can occur in the individual RAMS phases and what human factors activities might be associated with it. This leads to HOF being insufficiently considered or only coming into play in the later stages of projects, when adjustments are hardly feasible and costly. With the approach and model we have developed, we therefore offer concrete assistance for the systematic and verifiable inclusion of human factors throughout the system lifecycle.

The goal is to support the understanding of how to integrate human and organisational factors into railway system design from the perspectives of different stakeholders.

Method

The EN50126-1 V-cycle model was chosen as the basis for describing the steps involved in integrating human and organisational factors. This model describes 12 steps in the generic RAMS process from concept through to decommissioning. The first step was to define 12 human and organisational factors steps that aligned with the model, to describe the main human factors integration goals at each step of the RAMS process. These human and organisational factors steps were reviewed and refined with the UIC HOFWG in order to validate them. A further review and validation was then conducted by the ERJU PRAMS working group.

The 'classic' representation of the V-cycle as described in EN50126-1 is composed of 12 phases that can be divided into three major blocks (see Figure 1):

- Risk assessment and requirement development (i.e., '*Tender'* block, marked in red in Figure 1), incorporating phases 1 to 5.
- Implementation and demonstration of compliance with RAMS requirements (i.e. '*Design and Commissioning'* block, marked in orange in Figure 1), incorporating phases 5 to 10. Phases 8, 9 and 10 can be seen as an integration subphase.
- Operation, maintenance and decommissioning (i.e. '*Operation'* block, marked in blue in Figure 1), incorporating phases 11 and 12.

For all 12 steps of the V-cycle, HOF plays an important role in achieving the desired safety and performance and human factors specialists would expect to contribute at each stage. However, the HOF contribution is not explicit in the model and therefore may not be appreciated by system engineers, resulting in gaps in the application of HOF. Genuinely considering Human and Organisational Factors (HOF) from start to finish means applying a structured and practical approach to the integration of HOF. The solution we propose is an enrichment of the V-cycle model with HOF sub-steps and outputs at each stage.

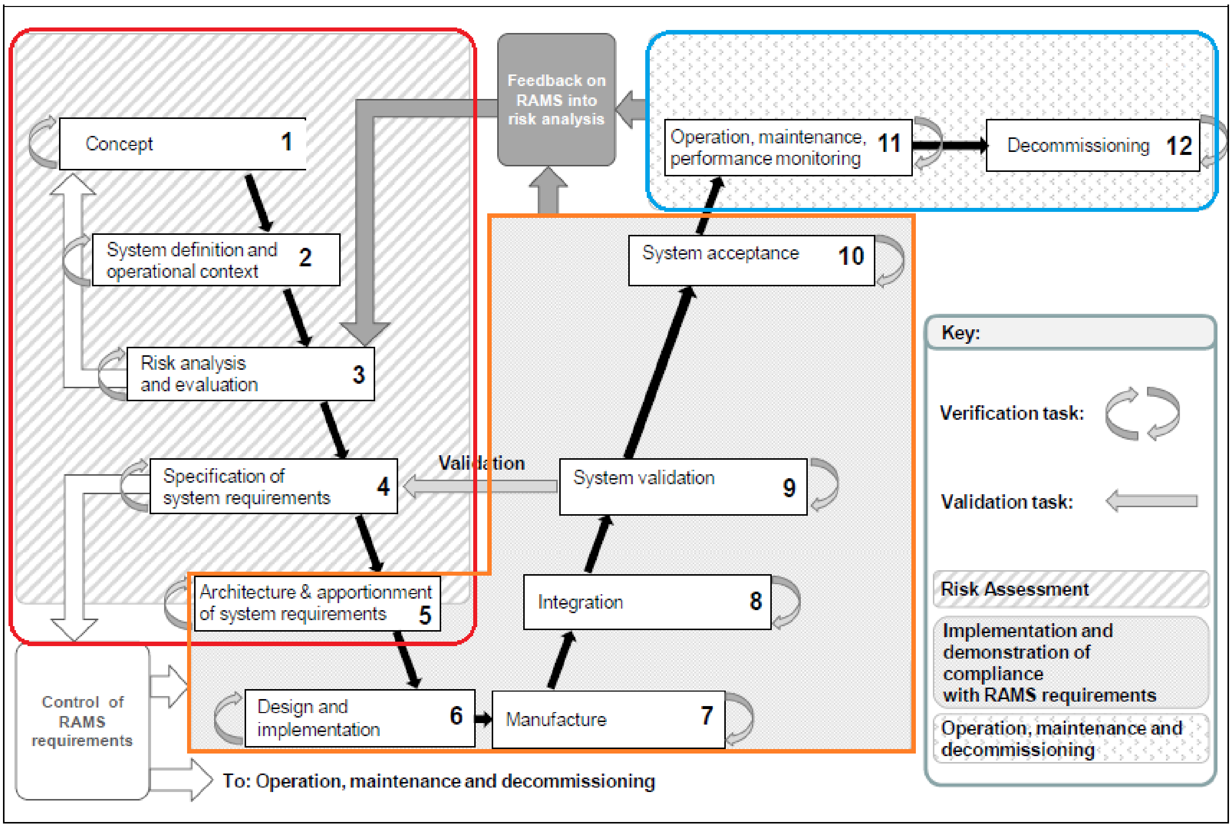


Figure 1: V-cycle model (Figure 7 from EN50126-1)

Solution

Railway organisations must have a standard process for implementing new systems, new trains, new safety equipment or major modifications (i.e., a change management process). This process is already supported from a systems engineering perspective by the EN50126-1 standard which lays out a generic process for the management of reliability and safety through system lifecycle design. Human factors is already addressed in a specific section of the standard which states that analysis of human factor is inherent within the systems approach taken by the standard. A non-exhaustive checklist of human influencing factors is included in the standard which include for example: human competence, the working environment, culture, human/machine communication, and interface design. However, the standard does not currently give any specific guidance in terms of *how* human factors can be integrated in the system lifecycle design, or even how these influencing factors can be applied in practice. This is in contrast to the integration of safety in the traditional RAM process, which was fully integrated in the EN50126 model in 2017 alongside the publication of EN50126-2 which provides guidance and methods on how to follow the Safety Management Process described in Part 1 of the same standard.

We propose that human and organisational factors can be integrated in the change management process by following the steps mapped to the EN50126-1 V-cycle model shown in Figure 1 below. This new solution proposes to use the stages that logically follow each other during the V-cycle (EN 50126), which is traditionally used as a reference to develop a new system whether they are safety equipment, new trains or any new system to set up within the company. HOF should be integrated throughout the V-cycle, i.e. in all the three phases. This is why we are suggesting a new approach to "implementing any new system within the organisation", with HOF being considered at the same level as the technical elements. This structure is used to describe the actions to be taken at each stage of the cycle to ensure that HOF is considered at the right time. The advantage of integrating HOF within the V-model is that HOF is systematically and verifiably incorporated, allowing the final system to be targeted at end user needs from the beginning and providing evidence to the regulator that HOF have been taken into account during the whole development process. The original 12 RAMS steps are shown in black, while the aligned HOF steps are shown in white.

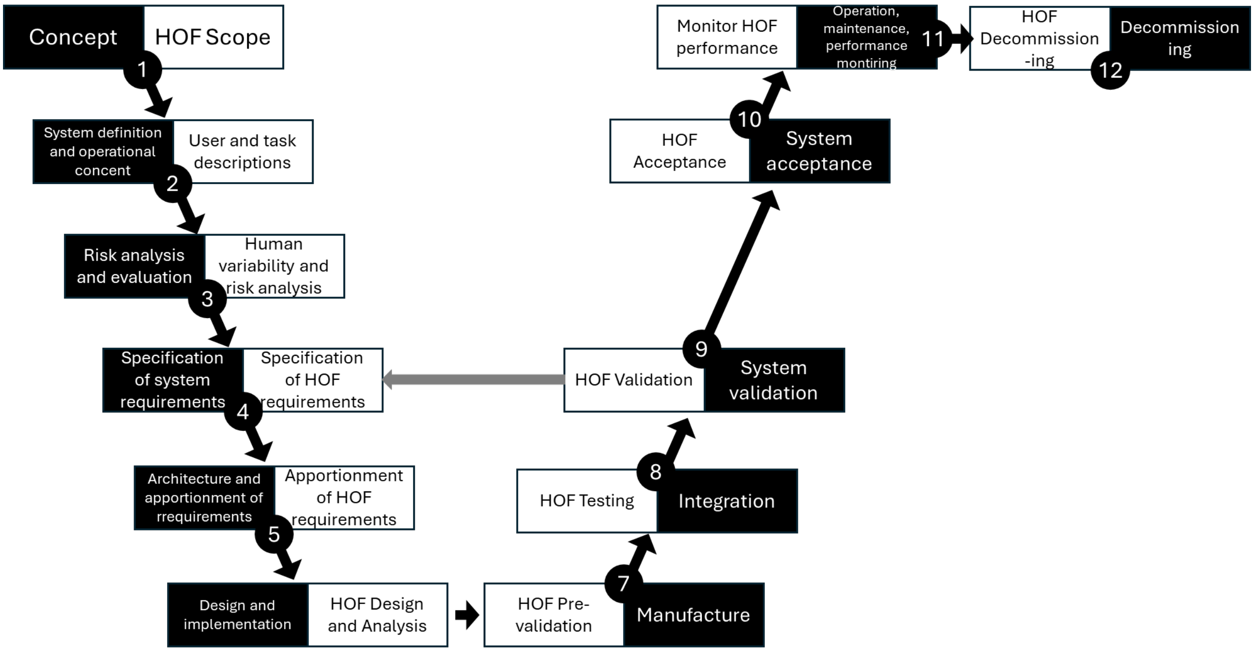


Figure 2: EN50126-1:2017 V-cycle model with additional human factors steps

The HOF steps are:

1. HOF Scope (Concept) – at this stage, the HOF Scope for the project should be set, considering the possible impact of the change on the end users and identifying high level areas of focus for the HOF work.
2. User and task descriptions (System definition and operational context) – at this stage, all end users of the system should be described in detail and a high-level description of their main goals and tasks provided. The major interfaces between the end users and technical system elements should be identified. The principal HOF goals should be defined to ensure that the users can perform their future roles effectively, drawing on the scope already defined. The HOF Assurance Plan should be produced at this stage.
3. Human performance analysis (Risk analysis and evaluation) – at this stage, HOF should be integrated with the overall risk assessment process to represent HOF hazards. Specific HOF risk assessments may be carried out into potential human performance risks. The HOF Issues Log (HOFIL) should be started at or before this stage.
4. Specification of HOF requirements (Specification of system requirements) – at this stage, HOF requirements on the end system should be developed and documented, including criteria for acceptance and means of compliance.
5. Apportionment of HOF requirements (Architecture and apportionment of system requirements) – at this stage, the HOF requirements should be apportioned to the relevant sub-systems. The HOF Assurance Plan should be updated.
6. HOF Design Analysis (Design and implementation) – at this stage, specific HOF assessments may be conducted to support the design process and provide evidence for closure of HOF issues. Human centred design is best suited to an iterative design process

where evolving designs can be tested with end users at different stages of design maturity and feedback used to improve the design. Wherever possible, this should be facilitated. HOF assessments should be conducted during the design process by the mean of simulation to determine the extent to which HOF goals are achieved and risks are mitigated. In certain cases, from an HOF perspective, application conditions may need to precisely state in which working conditions the system can operate safely.

7. HOF Pre-validation (Manufacture) – at this stage, the HOF focus is on preparing for validation of the final product. There may also be some HOF issues to be addressed relating to preparing for operations and maintenance.
8. HOF testing (Integration) – at this stage, the integration of the human component (end users) should be tested with the technical components they are required to interact with.
9. HOF Validation (System validation) – at this stage, the HOF requirements are validated against the final product and the HOF Assurance Report is produced documenting the HOF activities undertaken throughout the project, and their results. The acceptability of the design is finalised against the HOF requirements and with end users, and issues on the HOFIL should be closed out.
10. HOF acceptance (System acceptance) – at this stage, the HOF Assurance Report is reviewed and accepted by a reviewer with suitable HOF expertise.
11. Post-commissioning and in-service reviews (Operation, maintenance and performance monitoring) – post-commissioning reviews are performed with end users to close out any outstanding issues on the HOFIL and to identify any additional issues that may need to be resolved. In-service reviews may be carried out through the lifetime of the system to identify and resolve any operational or maintenance HOF issues that emerge over time.
12. Decommissioning – the HOF impact of planned decommissioning activities is assessed and a HOF plan put in place to manage any expected issues.

Table 1 below describes the HOF tasks for each of the 12 phases, aligned with the General Tasks from Table 1 of EN50126-1.

Table 1: HOF tasks for lifecycle phases 1-12

#	RAMS Phase	HOF Phase	General tasks	HOF Tasks
1	Concept	HOF Scope	Investigate scope, context, and purpose of system. Investigate the environment of the system.	Investigate the HOF impact of the system. Investigate HOF issues on similar systems. Identify relevant HOF standards. Define the HOF scope for the project.
2	System definition and operational context	User and task descriptions	Define the system and its mission profile. Define the system boundary. Define the scope of operational requirements. Establish the organisation.	Define the end users. Define high-level tasks of end users and their future roles. Identify end user interfaces with technical components/sub-systems. Define HOF goals. Establish the HOF Assurance Plan.
3	Risk analysis and evaluation	Human performance analysis	Identification and classification of hazards associated with the system.	Identify HOF risks and integrate them into the risk analysis.

				Conduct any necessary human performance analyses. Establish HOF Issues Log (HOFIL).
4	Specification of system requirements	Specification of HOF requirements	Specify system requirements.	Establish HOF requirements specification. Update HOFIL. Update HF Assurance Plan. Establish HOF requirements validation plan.
5	Architecture and apportionment of system requirements	Apportionment of HOF requirements	Define the system architecture. Identify the requirements for integration of pre-existing subsystems/components. Define acceptance criteria and processes for subsystems/components.	Allocate HOF requirements to sub-systems. Update HOFIL. Update HF Assurance Plan.
6	Design and implementation	HOF design analysis	Design subsystems/ components. Prepare operation and maintenance procedures. Define training measures for operation and maintenance. Define and establish manufacturing process for producing subsystems and components. Define and establish system integration process. Prepare installation and commissioning procedures.	Conduct iterative HOF assessments on the emerging design. Conduct user testing where possible. Provide feedback and recommendations on design improvements. Conduct HOF assessments on operation and maintenance procedures. Analyse if application conditions should be defined. Identify HOF training needs Update the HOFIL.
7	Manufacture	HOF pre-validation	Implement and operate manufacturing process.	Prepare HOF validation plan.. Update the HOFIL.
8	Integration	HOF testing	Integrate subsystems and components. Demonstrate system functionality. Test and analyse system. Arrange system support arrangements.	Conduct end user testing on the final product Update the HOFIL. Collect evidence that HOF requirements have been met.
9	System Validation	HOF validation	Establish validation report. Establish process for the acquisition and evaluation of operational and maintenance data.	Collect evidence that HOF requirements have been met. Collect evidence that HOF issues have been closed. Document results of all HOF activities. Produce HOF Assurance Report.

10	System acceptance	HOF Acceptance	Record an acceptance record. Verify the acceptance record.	Review of HOF validation activities and record of acceptance.
11	Operation, maintenance and performance monitoring	Post-commissioning and in-service reviews	Provide all information necessary to formulate plans/procedures for operation and maintenance. Implement operation and maintenance procedures. Record changes in the system configuration.	Conduct post-commissioning review. Conduct in-service reviews as necessary. Monitor HOF during operation and changes.
12	Decommission-ing	Decommission-ing	Establish decommissioning plan and related report.	Identify the HOF impact of decommissioning.

Conclusions and next steps

The model presented in Figure 1 describes a generic process for the integration of HOF in change management of railway systems, aligned with the V-cycle in EN 50126-1. It provides a closer and more systematic link between the HOF knowledge of the industry and practice in the transport sector, which is not limited to modelling and designing trains, but to all relevant system changes in the rail industry, based on the V-cycle used in system engineering. This model is the starting point for additional work to further specify the human and organisational factors steps and produce HF process requirements to support railway organisations engaging suppliers to integrate human and organisational factors during the system lifecycle. The UIC HOFWG, in collaboration with the PRAMS group, will focus next on producing guidance on the application of each step. The eventual output should be a standardised approach for integrating human and organisational factors into the railway systems lifecycle.

Acknowledgements

The authors wish to acknowledge the input of the full UIC HOFWG in developing the model describes in this paper, and the ERJU members: Frits Neuteboom (NS), Maarten Devries (NS), and Paolo Stoico (Mermec Group).

References

Balfe, N. 2023. H-FIT: Assessing the human factors impact of proposed changes to the railway. In Proceedings of the 2023 UK Ergonomics and Human Factors Conference. Birmingham: CIEHF.

EN50126-1, 2017. Railway applications – The specification and demonstration of reliability, availability, maintainability, and safety (RAMS). Part 1: Generic RAMS process. Brussels: Cenelec.

EN50126-2, 2017. Railway applications - The specification and demonstration of reliability, availability, maintainability, and safety (RAMS). Part 2: Systems Approach to Safety. Brussels: Cenelec.

ERA, 2023. Guidance: Human and organisational factors integration in change management. Valenciennes: ERA.

Eurocontrol. 2011. Support material for Human Factors Case application. Brussels: Eurocontrol.

ISO9241-210, 2019. Ergonomics of human-system interaction. Part 210: Human-centred design for interactive systems. Geneva: ISO.

Understanding work-as-done: Lessons learned in a case study on fatigue risk management

Anisha Tailor, Anna Vereker & Tom Hyatt

Rail Safety and Standards Board, United Kingdom

SUMMARY

This paper presents a case study of a workshop to understand rail track workers' experiences of organisational fatigue risk controls. The design and facilitation of the workshop drew on key skills, attitudes and practices for learning about work-as-done. This paper describes the successes, challenges and lessons learned associated with practically embedding these into the workshop.

KEYWORDS

Fatigue, safety culture, rail

Background and aim

A comprehensive fatigue risk management system involves a partnership between management and the workforce, to continuously monitor and manage fatigue risk (Gander, Hartley, Powell, Cabon, Hitchcock, Mills et al., 2011). Rail infrastructure companies employ fatigue risk controls, such as the provision of hotel accommodation, to mitigate the risks associated with night working and travelling long distances between site and home. However, companies face challenges in understanding front line workers' knowledge, take-up and experience of using such controls.

Understanding whether and why workers do or do not use available fatigue controls involves learning about 'work-as-done'. Shorrock (2020) set out essential skills and attitudes for learning from everyday work. The three skills are: 'ask good questions'; 'listen well', and 'take multiple perspectives'. The three attitudes are: willingness; curiosity, and humility.

Recent RSSB research also identified ways of engaging with the workforce that are perceived to be particularly effective for learning about work-as-done (Tailor, McCulloch, & Lonergan, 2024):

- Put people at ease and be relatable,
- Treat people as experts in work-as-done and avoid sounding like you are testing them,
- Allow discussions to deviate from pre-determined lines of enquiry.

RSSB Human Factors Specialists facilitated two, simultaneous small-scale workshops with 16 rail track workers, who worked directly or indirectly for one infrastructure company. This was to help understand the 'work-as-done' aspect of the company's fatigue risk management practices. The facilitators sought to incorporate the practices, skills and attributes identified by Shorrock and Tailor et al into the design and facilitation of the workshop. This paper presents successes, challenges and lessons learned in putting these approaches into practice.

Applying the skills, attitudes and practices: Successes and challenges

Facilitators created a briefing note for the company to send to prospective participants, rather than relying on managers to relay this information on their behalf. This allowed them to explain the purpose and format of the workshop and reassure participants in their own words. A semi-structured question guide was developed. It started with open questions that aimed to empathise with participants (e.g. what impact does night shift work have on you personally?), before later focusing in on company fatigue controls. The questions about fatigue controls started broadly (e.g. how does the company help you manage fatigue?) to allow lines of enquiry to evolve and to avoid suggesting that the facilitators were seeking a 'right answer'. At the end, participants were asked what they thought would help with managing fatigue.

On the day, participants were welcomed to the workshop in a relaxed and friendly tone as they entered. The facilitators purposely wore casual clothes (such as jeans) rather than usual office wear to help put participants at ease. In the opening brief, facilitators acknowledged that they were novices in front line track work and that the participants were the experts. To focus on the discussion, technology was not used during the workshop. Note-taking was via flipcharts and pens and paper. This allowed facilitators to be transparent about what was being written. Active listening techniques were used, such as reflecting and summarising what participants said. Facilitators were careful to avoid correcting participants, for example if an individual said something contrary to scientific evidence about fatigue, or incorrectly described an operational procedure. Discussions about adjacent but not directly relevant topics, such as operational processes, were allowed to continue for some time before moving on. This was to avoid stifling discussions. Facilitators asked clarifying questions when necessary. This involved being comfortable with drawing attention to gaps in their operational knowledge.

The workshop took place in the company's head office between 21:00 and 23:00, before a night shift. Ideally, facilitators would have engaged with participants in their normal working environment, to meet them 'in their territory'. However, there are limited welfare facilities for track workers, so it was deemed neither safe nor practical to do this. An online workshop had been considered but it was concluded that this would compromise the facilitators' ability to build rapport with a hard-to-reach group. During the workshop, facilitators learned that the timing and location of the workshop had inadvertently interfered with participants' usual preparation for a night shift. Participants had had to make an extra car journey from home to the office, before travelling to their work sites. For most participants, the work sites were not located close to the office or home. Participants had therefore had to eat their evening meal and leave home earlier than usual. An additional challenge was that focusing on listening rather than taking high-fidelity notes presented some challenges when writing up the findings, as this placed greater reliance on facilitators' memory.

Discussion

Despite the challenges discussed, participants were open, and the workshop provided novel insights and feasible suggestions to the company. Ideas for improvement included adjustments to shift patterns and better methods of communicating fatigue-related material. In future, we would consider a different compromise: carrying out an online workshop to prioritise participants' schedules, accepting a potential reduction in rapport and group dynamics.

The skills, attributes and practices identified in the work by Shorrock and Tailor et al may seem like 'common sense' to the Human Factors community, where a natural curiosity about others' work and experiences may have led individuals to this profession. Nevertheless, this work draws attention to the constraints and trade-offs faced by Human Factors professionals – our own 'work-as-done' reality. It also highlights the need to carefully weigh up the impact of decisions on participants, and to continually reflect on attitudes and hone facilitation skills. These are essential to Human Factors professionals, present and future.

References

Gander, P., Hartley, L., Powell, D., Cabon, P., Hitchcock, E., Mills, A., & Popkin, S. (2011). Fatigue risk management: Organizational factors at the regulatory and industry/company level. Accident Analysis & Prevention, 43(2), 573-590.

Shorrock, S. (2020). Why learn from everyday work? Hindsight Magazine, 31, 8-9. Brussels, Belgium: Eurocontrol.

Tailor, A., McCulloch, O., & Lonergan, J. (2024). Enhancing organisational learning about how track worker safety is managed. Rail Safety and Standards Board.

Verification and Application of Workwear-Integrated Sensors for Ergonomic Injury Risk Assessment

Arun Nandakumar[1], Manal Elhamri[1], Zain Shah[1], Kailash Manohara Selvan[2], Stephen Womack[3], Simon Carpenter[4]

[1]Imperial College School of Medicine, UK [2]SpatialCortex Technology Ltd, UK, [3]Amey Infrastructure Wales, UK. [4]Transport for Wales, UK

SUMMARY

Risk assessments are a central part in Work-related Musculoskeletal Disorder (MSD) risk management. Wearable sensors integrated into workwear present a promising approach for industrial ergonomic risk assessment This paper has three objectives: first, we assess the accuracy and efficacy of a workwear-integrated sensor system in estimating the MSD risks comparing it to an optical measurement technique, secondly, we outline approaches for using the system for continuous MSD risk assessment using industry standard methods and a Cumulative Damage (CD) exposure metric based on Fatigue Failure Theory (FFT) and third, we demonstrate feasibility by applying the system and approach for assessing 3 manual-handling tasks within the rail industry. Results confirm that the system, and approach can enable a proactive and data-driven approach to MSD management. Key considerations necessary for the wider adoption are also presented.

KEYWORDS

Musculoskeletal Disorders, risk assessment, sensors, rail industry

Introduction

Musculoskeletal Disorders (MSDs) are a significant contributor to workplace sickness absence. Assessing risks and reducing them to be As Low As Reasonably Practicable (ALARP) is central to any MSD risk management. Most of the current risk assessment tools rely on snapshot assessments of single postures, often those believed to be hazardous or problematic (A. D. J. 2002). These risk assessments primarily rely on visual inspections, which are subjective and suffer from accuracy issues. Additionally, the lack of continuous assessment may prevent these methods from capturing risks across the entire work shift and accurately reflecting staff behaviours. MSDs are one of the leading causes of absence for the rail workforce (RSSB, 2024) and failure to conduct and apply the findings of a suitable and sufficient risk assessment has been reported as underlying cause (ORR, 2019). Recent advances in movement sensor technology and algorithms allow workwear-integrated sensors to track ergonomic risks. This study reports accuracy verifications conducted on the workwear-integrated sensor system, approaches for using workwear-integrated sensor systems to assess MSD risks in a continual and cumulative basis and evaluates the feasibility on 3 rail -sector manual tasks.

Accuracy verification of the work wear integrated sensor system

The system features 9-11 sensors measuring movement (SpatialCortex MOVA) embedded into standard workwear at key locations (limbs, torso and lower back) and a wearable data logger.

Accuracy of the sensor system was verified against a novel optical measurement system. The reliability of the optical system was measured initially. Accuracy verification of the sensor system relied on markers placed at various locations of participants adopting a range of poses, sensor data and camera images captured concurrently, and each data frame analysed to extract, joint angles, horizontal distance from lower back and vertical distance from the ground. The study also considered the effect of clothing, participant variability and sensor attachment methods on system accuracy.

Camera-based optical measurement technique

A novel optical measurement technique was developed using readily available materials, including reflective markers, a standard camera, and a free image analysis software package. The method aims to provide an accessible and practical approach for benchmarking sensor system accuracy in various work environments. The measurement process involved placing at least two markers (separated by a set distance) per sensor unit's front face, attaching the sensors to the participant, and capturing 2D images alongside sensor data while the participant adopted a series of predefined poses. Data acquisition was synchronised via a Python-driven script using the MOVA software suite. Three additional reflective markers defined the origin, horizontal, and vertical datums of the measurement environment. A 2D digital image processor (PlotDigitizer package) extracted marker coordinates based on these datum references. The orientation of each sensor was determined from the coordinates of its marker pair, enabling joint angle calculations when sensors were placed on multiple body segments (e.g., upper arm and hand). Segment lengths measured and the orientation from marker data facilitated the measurement of horizontal and vertical distances. While previous research (Aydin, 2021) reported high reliability for the software tool and approach in extracting scientific data from 2D images, the manual nature of the process warranted additional Repeatability and Reproducibility Assessments (RRA) to evaluate the method's reliability comprehensively.

RRA were conducted using three selected frames (Figure 1, a-c) and three operators, each manually extracting sensor segment orientations with reflective markers using the PlotDigitizer package, performing three repetitions per operator. For repeatability, the Standard Deviation (SD_r), Coefficient of Variation (CV%), and Intra-Class Correlation (ICC) ranged from 1.07° to 5.38°, 1.2% to 5.6%, and 0.85 to 0.98, respectively.

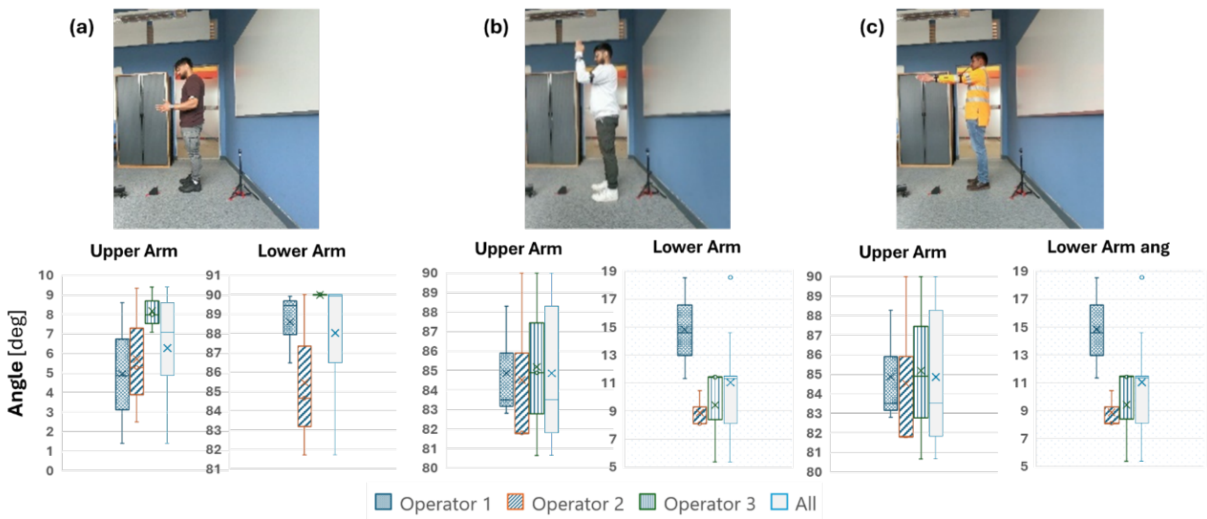


Figure 1: Repeatability and Reproducibility assessment of the camera-based optical technique

For reproducibility, the Standard Deviation (SD_o), Inter-Operator Variability (IOV%), and Intra-Class Correlation (ICC) ranged from 0.44° to 2.20°, 0.5% to 2.5%, and 0.89 to 0.98, respectively.

The optical measurement system demonstrates good reliability, with results consistent within the reported bands.

Angular and spatial accuracy verification

The study focused on measuring the elbow flexion joint angle, where the orientation of the upper and lower arm was measured using both the sensor and optical techniques. While the approach is transferable to other joint groups, this report presents findings exclusively from the elbow joint measurements.

Measurements were obtained from six predefined postures, repeated across three participants to capture person-to-person variability. To evaluate the influence of clothing and sensor attachment methods, the following conditions were tested: (a) short-sleeve shirt with sensors attached directly to the skin using Velcro straps, (b) long-sleeve shirt with sensors attached over a fabric layer using Velcro straps, (c) orange high-visibility long-sleeve top with sensors attached similarly to condition b, and (d) green high-visibility jacket with sensors inserted into specialized pockets containing elasticated segments for sensor stability.

The results (Figure 2) indicated that the mean absolute error (MAE) for upper arm orientation ranged from 4.6° to 6.1° across individuals, with an overall MAE of 5.1°, standard deviation of 4.7°, and a maximum error of 15.8°. For lower arm orientation, the MAE ranged from 7.3° to 10.1°, with an overall MAE of 8.7°, standard deviation of 3.8°, and a maximum error of 14.8°. These angular errors translated to spatial errors of 3.5 cm MAE, 3.0 cm standard deviation, and 9.0 cm maximum error for horizontal distance from the lower back, and 4.2 cm MAE, 4.3 cm standard deviation, and 13.0 cm maximum error for vertical distance from the ground.

Analysis of the effect of clothing and attachment methods (Figure 3) revealed minimal impact compared to person-to-person variability. Measurements were consistent regardless of whether sensors were directly attached to the skin or placed within pockets over elastic segments.

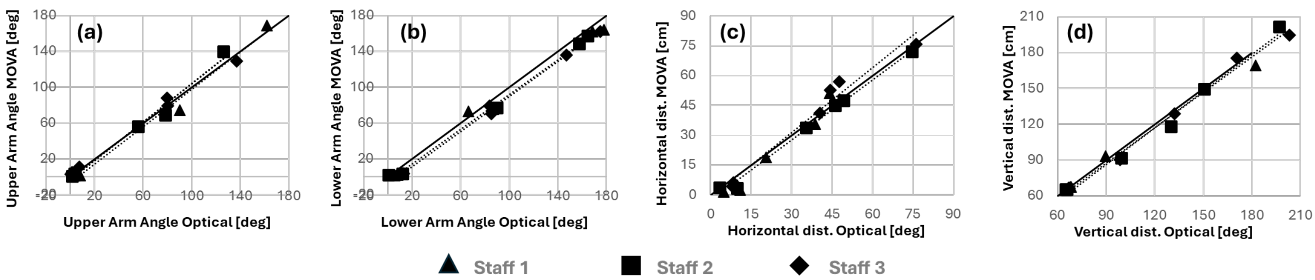


Figure 2: Accuracy of workwear integrated sensor system: Staff-to-Staff variability

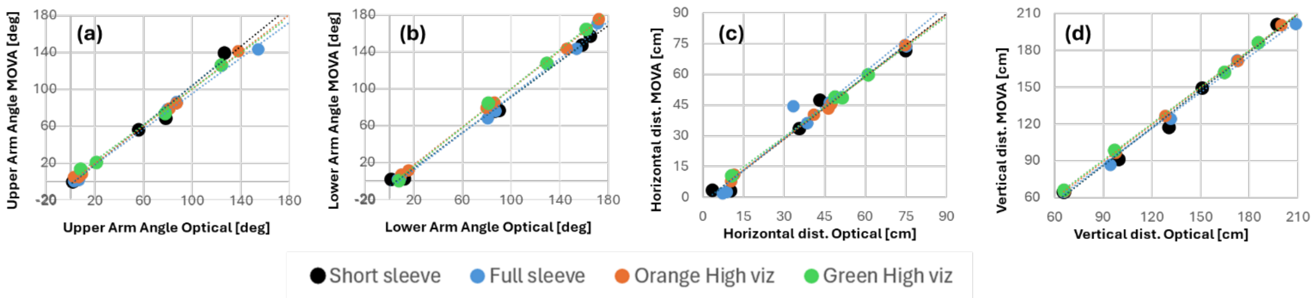


Figure 3: Accuracy of workwear integrated sensor system: Effect of clothing

Sensitivity Analysis of lower back biomechanical parameter estimation

To quantify the impact angular and spatial measurement errors on estimates of key biomechanical parameters; L5/S1 Moment ($M_{L5/S1}$) and L5/S1 Spinal compression (Fc), a Sensitivity Analysis (SA) was conducted. These parameters were calculated following the approach presented by Chaffin, D.B. (2006),

$$M_{L5/S1} = E.F_M = (b.mg_{BW} + h.mg_{Load}) \quad \text{(Equation 1)}$$

$$Fc = -Cos(\alpha)[mg_{BW} + mg_{Load}] - F_M \quad \text{(Equation 2)}$$

where E is the erector spinae moment arm, F_M is the effective reactor spinae muscle force, b is the upper body centre of gravity - lower back horizontal distance, h is the load – lower back horizontal distance, mg_{BW} is the upper body weight, mg_{Load} is the load weight and α is the sacral cutting plane angle with the horizontal.

A Monte Carlo based SA was conducted, assuming a baseline posture depicted in Figure 4(a), a load weight of 15 kg, and SD for the upper and lower arm orientations derived from the accuracy measurements. A ±15% bound was able to explain the observed variability in $M_{L5/S1}$ and Fc (Figure 4, b-c). Conservatively, a 15% margin can be applied to any estimates of L5/S1 compression force derived from the sensor system's data. It is important to note that this analysis does not account for potential errors in torso angle measurement. Therefore, the bounds should be increased further for a more conservative and safer estimation.

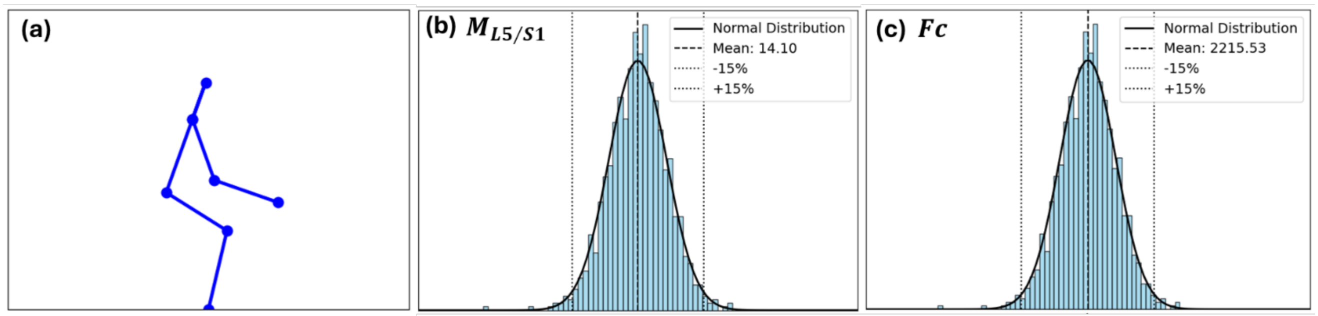


Figure 4: SA of biomechanical parameters; $M_{L5/S1}$ and Fc, using inputs from workwear-integrated sensor system

All results presented here were based on data collected within two hours of participants wearing the sensor system after calibration. A critical consideration for future work is to investigate the impact of prolonged sensor use, such as an 8-hour work shift, where sensor movement due to physical activity could introduce additional variability into the measurements.

Approaches for MSD risk assessment with data from workwear-integrated sensors

Two methods were chosen for MSD risk assessment; Health and Safety Executive's (HSE) Manual Handling Assessment Charts (MAC) and a method based on Fatigue Failure Theory (FFT) for continual and cumulative risk exposure assessment respectively.

Continual MSD risk assessment using HSE MAC risk factors

The MAC tool evaluates risks in lifting, lowering, carrying, and team manual handling tasks using a numerical and color-coded scoring system to identify high-risk activities based on established risk factors. The V-MAC tool extends this capability by assessing tasks with variable load weights. The MAC tool was selected due to its ease of use and suitability for manual handling risk assessment

(Pinder, 2002). Validation studies identified 'Horizontal distance' as a significant predictor of work-related Lower Back Pain (Pinder, 2014).

Cumulative MSD risk assessment using an improved Fatigue Failure Theory (FFT) based model

Recent evidence suggests that MSDs such as Lower Back Pain may be the due to a fatigue failure process (Gallagher and Schall, 2016) accumulating during repeated loading cycles, which allows predicting a Cumulative Damage (CD) metric. Gallagher, S et al., 2017, presents the Lifting Fatigue Failure Tool (LiFFT), with three inputs (load weight, peak horizontal distance from spine to load, and repetition) and provides validation against two existing epidemiological databases; mono-task jobs (Marras et al., 1993; Zurada et al., 1997), variable loading jobs (Sesek,1999), the model explained deviance within 92% and 72-95% in the database respectively.

The LiFFT model calculates CD by computing the Peak Load Moment (PLM), the product of the object's weight and its horizontal distance from the spine. It then estimates spinal compression and calculates damage per cycle (DPC), comparing this compression to the average spine's compressive strength (6 kN; Jager & Luttmann, 1991). DPC is multiplied by the task repetitions, and a CD threshold is applied to determine risk. The model's simplicity, minimal input requirements, and ability to estimate daily CD dose across multiple subtasks are key advantages.

However, the accuracy of the compression force (Fc) calculation is critical. The simplified Fc calculation (Equation 3) omits factors like torso posture and upper body weight, which are significant and considered in Equations 1 and 2. This simplification can cause errors, especially when loads are held away from the body. To address this, a scaling factor is applied to the CD estimate, though this remains a coarse adjustment.

$$Fc_{LiFFT} = \frac{1}{E}\ h.mg_{Load} \qquad \text{(Equation 3)}$$

The workwear-integrated sensor system provides a key advantage by enabling complete, time-dependent Fc calculations based on sensor estimates when the task's peak load weight is known. This allows for real-time tracking of cumulative damage (CD) throughout the work shift.

Feasibility assessment: application to rail industry manual-handling tasks

The workwear-integrated sensor system was evaluated in three rail industry manual-handling tasks:

1. Manual-material transfer**:** Staff moved construction materials, such as concrete bags, from access points to platforms for loading onto rail trolleys. Certain access points presented logistical challenges, making mechanized transfer impractical in all cases.
2. Rail ramp deployment**:** Onboard hosts, station staff, and passenger assistance staff regularly deploy ramps of varying designs. Heavier ramps are equipped with wheels to reduce manual lifting, while lighter (~10 kg) ramps are often carried. Ramp storage locations, onboard or on platforms, further influence carrying demands.
3. Manual vegetation clearance**:** Maintenance staff manually cleared vegetation along the rail infrastructure using chainsaws. The task involved repetitive cutting at ground level and manual debris removal to maintain safe clearance from the running line.

These tasks were selected to demonstrate the system's applicability for ergonomic risk assessments across diverse job roles in the rail sector. Future task selection may use alternative metrics, such as historical injury data or lost-time incident records.

System setup: Participants wore standard rail-specific workwear with sensors attached using Velcro straps, requiring no garment modifications. Data was logged on a datalogging device kept in the worker's pocket. The system was calibrated with static and dynamic poses in under five minutes on average. Basic anthropometric data was collected beforehand to personalize posture risk thresholds for each participant.

Continual MSD risk assessment using HSE MAC risk factors

The HSE MAC assessment of manual lifting and handling tasks considers various risk factors, including A. Load weight/frequency, B. Horizontal distance from the lower back, C. Vertical lift distance, D. Torso posture, E. Postural constraints and additional factors include carry distance, grip, floor surface, obstacles, and environmental conditions.

The workwear-integrated sensor system can directly measure risk factors D, B, and C, as shown in Figure 5 (b)., for manual material transfer tasks. The load handled was concrete bags each weighing 20Kg. These risk factors are plotted continuously over time. The torso posture (risk factor D) chart tracks torso flexion and sideways angles, showing spikes during lifting or lowering events. The flexion angles range from 30° to 70°, with the worst-case posture during lowering showing the largest flexion and horizontal load distance (B) extending into the red zone. This data indicates opportunities to train staff on improving posture.

The vertical lift distance (C) indicates that lifting starts above elbow height, while lowering often extends below knee level, compromising posture. The torso posture chart (D) can also help infer repetition and frequency, data shows repetitive lifting and carrying in bursts, with high-intensity transfer phases followed by lower-intensity periods. The maximum lift frequency during a burst can elevate the risk associated with this factor. A conservative approach using the worst-case posture and lift frequency rates the task as moderate-to-high risk. Additionally, gait patterns estimating the number of steps taken. The median carry distance was found to be 6.6 meters, aligning closely with the actual work environment value of 6 meters.

This data provides valuable insights for Health and Safety (H&S) duty holders to evaluate control measures, such as eliminating manual transfer where feasible, substituting lighter loads, implementing mechanized transfer where practical, training staff on proper lifting techniques, improving job planning to reduce postural constraints, using job rotations to minimize exposure to high-risk tasks.

Data gathered from the rail ramp deployment task involved staff walking to the train, accessing the ramp stored onboard, carrying it to the entrance, and deploying it. The data also covered the reverse operation of stowing the ramp away. The analysis revealed during ramp deployment staff spent extended time in high flexion, with hands extended away from the lower back and below knee level. The data also suggested that finer adjustments may be necessary to ensure proper ramp deployment. Such patterns can help identify opportunities for improving staff training, ramp design, and addressing ergonomic challenges related to ramp maintenance and deployment.

Similar data gathered from manual vegetation clearance tasks showed that staff engaged in ground-level cutting adopted various postures, often holding those positions for extended periods. While the load weight (chain saw) may not be significant, the prolonged postures contribute to ergonomic risks. This data can guide H&S stakeholders in evaluating appropriate controls, following the hierarchy of controls, such as mechanization where feasible, providing alternative tools to maintain good posture, offering task-specific training, promoting postural best practices, and implementing job rotations to minimize prolonged exposure to risky postures.

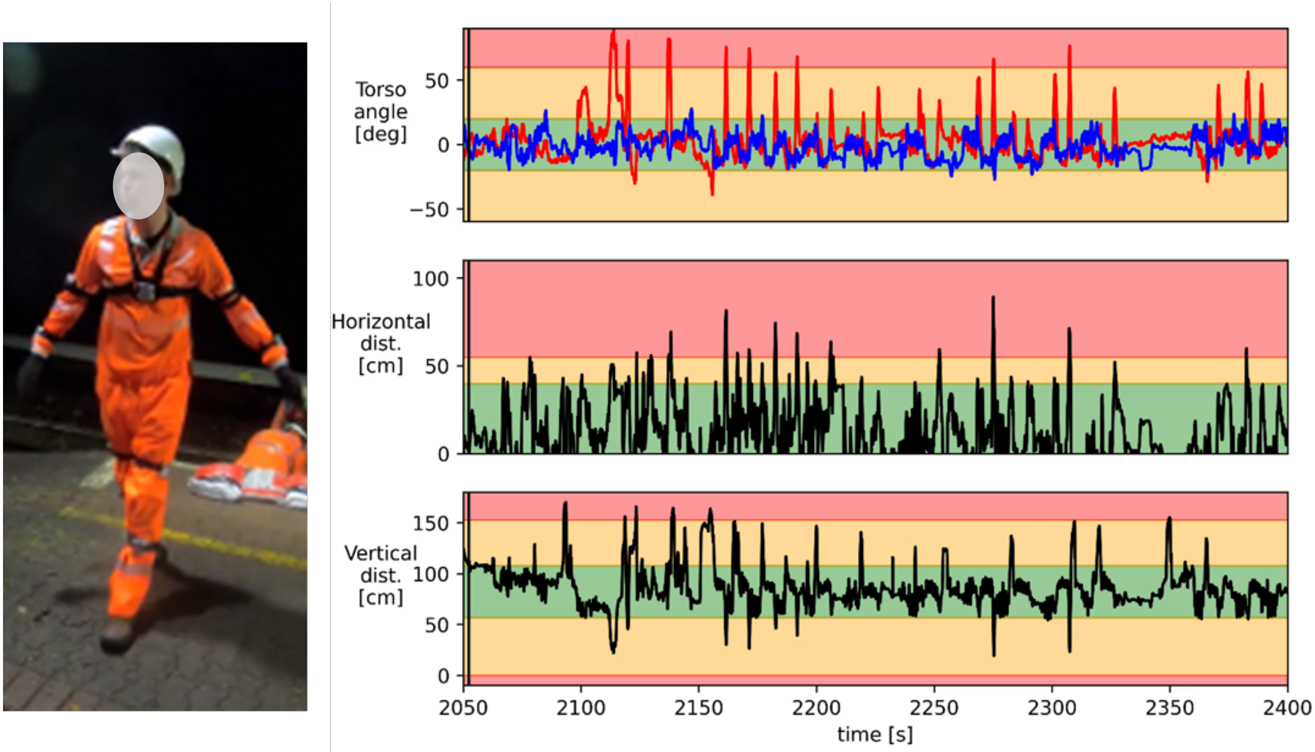


Figure 5: (a) staff with workwear-integrated sensor conducting manual-material transfer task in rail infrastructure (b) HSE MAC risk factors (D, B, C) from workwear integrated sensor system

The HSE MAC tool is effective for manual handling risk assessment due to its simplicity and intuitive RAG (Red-Amber-Green) colour coding. However, it relies on subjective, snapshot assessments during visual inspections, which capture risks only at a single point in time and require significant H&S officer involvement, limiting scalability. The workwear-integrated sensor system overcomes these limitations by continuously capturing posture risk data throughout the work shift. This data, when analysed alongside load weights, reveals high-risk manual handling activities and highlights discrepancies between 'task as planned' and 'task as conducted.' Continuous, objective risk monitoring enables more accurate, data-driven interventions. Insights support targeted mitigations following the hierarchy of controls, such as elimination and substitution where feasible, task redesign, equipment modifications, and training, for maintaining manual handling risks at ALARP levels.

Cumulative MSD risk assessment using an improved FFT model

CD scores derived using L5/S1 compression metrics, highlighted material transfer and vegetation clearance had high risk due to repetitive bending and sustained postures. These metrics can enable designing work shifts and task planning in a data-driven manner to reduce risk exposure and considering individual staff circumstances during job planning like managing return-to-work.

The Borg Rating of Perceived Exertion (RPE) is a psychophysiological scale designed to assess an individual's subjective perception of physical exertion during activity. The RPE scale is widely used in ergonomic assessments to evaluate physical workload, muscle fatigue, and task demands, particularly in manual handling tasks where it can complement objective measures (Borg, 1998).

Staff feedback showed a strong correlation between CD scores and perceived exertion (RPE) which reinforces the reliability of the risk ranking. The sensors, when integrated into well-fitted workwear, were reported as comfortable and non-intrusive, with no impact on task performance.

Table 1: Cumulative Damage (CD) score based on FFT model

Task	mg_{Load} [N]	Cumulative Damage (CD)	Borg RPE rating
Manual material transfer	196.2	0.01904	14
Ramp deployment	98.1	0.00021	10
Vegetation clearance	78.5	0.00919	13

Conclusion

MSDs are a leading cause of workplace absence, especially within the rail sector. Effective MSD risk management requires robust risk assessments, yet failures to conduct or apply such assessments are commonly reported. Traditional risk assessment methods rely on subjective, single-moment observations, often missing evolving risks. This study introduces workwear-integrated sensors for continuous, objective ergonomic risk monitoring without requiring visual inspections.

- Firstly, the workwear integrated sensor system's accuracy was verified using a camera-based optical measurement technique built with readily available hardware and software. Repeatability and reproducibility analyses confirmed acceptable reliability. Angular errors were measured within 5.1° (upper arm) and 8.7° (lower arm), translating to spatial distance errors of 3.5 cm MAE horizontal distance (max 9 cm) and 4.2 cm MAE vertical distance (max 13 cm). These results demonstrate the system's capability to reliably classify ergonomic risk boundaries.
- Secondly, the system's feasibility for continuous MSD risk assessment using HSE MAC guidance and cumulative risk estimation using a fatigue failure-based model was explored.
- Thirdly, the system was deployed in three distinct manual-handling tasks in the rail sector, confirming its practical applicability in real-world environments. The workwear-integrated sensor system enables proactive, data-driven risk mitigation. Insights can guide targeted interventions, such as task redesign, equipment modifications, and tailored training, work shift redesign supporting the reduction of manual handling risks to ALARP levels.

The findings support the potential of workwear-integrated sensors as a scalable, quantitative tool for MSD risk assessment in complex, dynamic work environments.

Outlook

Workwear-integrated sensors offer significant potential for MSD risk assessment, but broader adoption requires further development. Technologically, long-term validation over entire work shifts is necessary to assess sensor durability, accuracy over time, and practical considerations such as battery life, data transmission, and ease of deployment. Further refinements in sensor algorithms and integration with existing digital safety infrastructures could enhance usability. From a workforce perspective, ensuring comfort, usability, and minimal disruption to daily tasks is essential for staff acceptance. Transparent data protection policies must address privacy concerns, particularly around individual performance monitoring and data ownership. At the organizational level, seamless integration with existing H&S protocols is crucial. Demonstrating tangible benefits—such as injury reduction, improved work planning, and cost savings—will strengthen the business case for adoption. Additionally, cultural factors play a vital role; fostering a workplace environment that prioritizes ergonomic risk management and proactive safety interventions is necessary for sustained implementation. Collaboration between industry stakeholders, regulatory bodies, and technology providers can drive standardization and ensure that sensor-based risk assessments align with evolving occupational health frameworks. With further validation and integration, workwear-integrated sensors have the potential to revolutionize MSD risk management, offering real-time, data-driven insights that enable targeted interventions and long-term workforce protection.

Acknowledgement

The authors thank Transport for Wales and Amey Infrastructure Wales for supporting the industrial use cases and granting permission to submit this article.

References

Aydin, O., & Yassikaya, M. Y. (2021). *Validity and Reliability Analysis of the PlotDigitizer Software Program for Data Extraction from Single-Case Graphs*. *Perspectives on Behavior Science*, 45(1), 239–257. https://doi.org/10.1007/s40614-021-00284-0

Borg, G. (1998). *Borg's Perceived Exertion and Pain Scales*. Champaign, IL: Human Kinetics.

Chaffin, D. B., Andersson, G. B. J., & Martin, B. J. (2006). *Occupational biomechanics* (4th ed.). Hoboken, NJ: Wiley.

Gallagher, S., & Schall Jr., M. C. (2016). Musculoskeletal disorders as a fatigue failure process: Evidence, implications, and research needs. *Ergonomics*, 1–15.

Gallagher, S., Sesek, R. F., Schall, M. C., & Huangfu, R. (2017). Development and validation of an easy-to-use risk assessment tool for cumulative low back loading: The Lifting Fatigue Failure Tool (LiFFT). *Applied Ergonomics*, 63, 142–150.

Gordon, C. C., Churchill, T., Clauser, C. E., Bradtmiller, B., McConville, J. T., Tebbetts, I., & Walker, R. A. (2014). *Anthropometric Survey of U.S. Army Personnel: Methods and Summary Statistics (ANSUR II)*. Technical Report NATICK/TR-15/007. U.S. Army.

Jager, M., & Luttmann, A. (1991). Compressive strength of lumbar spine elements related to age, gender, and other influencing factors. In *Electromyographical Kinesiology*, 291–294.

Marras, W. S., Lavender, S. A., Leurgans, S. E., Rajulu, S. L., Allread, W. G., Fathallah, F. A., & Ferguson, S. A. (1993). The role of dynamic three-dimensional trunk motion in occupationally related low back disorders: The effects of workplace factors, trunk position, and trunk motion characteristics on risk of injury. *Spine*, 18, 617–628.

Office of Rail and Road. (2019). *Closing the Gap on Health: Review of Health Risk Management in the Rail Industry (2014–2019)*. Office of Rail and Road, United Kingdom.

Pinder, A. D. J. (2002). *Benchmarking of the Manual Handling Assessment Charts (MAC)* (HSL/2002/31). Health and Safety Laboratory.

Pinder, A. D. J., & Frost, G. A. (2014). *Validation of the HSE Manual Handling Assessment Charts as Predictors of Work-Related Low Back Pain*. Health and Safety Laboratory, Buxton, Derbyshire, UK. Prepared for the Health and Safety Executive.

PlotDigitizer. (2025, February 15). *Extract Data from Images*. https://plotdigitizer.com

Rail Safety and Standards Board (RSSB). (2024). *Managing Musculoskeletal Health in the Rail Industry: Roadmap for Reducing the Risk of Musculoskeletal Disorders (MSDs)*.

Sesek, R. F. (1999). *Evaluation and refinement of ergonomic survey tools to evaluate worker risk of cumulative trauma disorders* (Dissertation).

University of Michigan Center for Ergonomics. (2024). *3D Static Strength Prediction Program (3DSSPP) User's Manual*. University of Michigan, Ann Arbor, MI.

Zurada, J., Karwowski, W., & Marras, W. S. (1997). A neural network-based system for classification of industrial jobs with respect to risk of low back disorders due to workplace design. *Applied Ergonomics*, 28, 49–58.

Exploring responses to public transport disruption when travelling with young children

Cara Tyrrell[1], Katie Parnell[1] & Katie Plant[1]

[1]Southampton University, UK

SUMMARY

The aim of this study was to understand how adults respond to public transport (PT) disruption when travelling with young children. A series of semi-structured interviews were conducted based on a hypothetical scenario of a 45-minute delay. Analysis of responses revealed distinct differences in how adults respond when travelling accompanied versus alone. Considering the findings a series of recommendations are made which will improve experience of PT when travelling with young children and potentially encourage greater use of PT networks.

KEYWORDS

Public transport, Disruption, Encumbered, Children

Introduction

Public transport (PT) contributes to improved environmental, economic and health conditions in urban environments (Sun & Cui, 2018) and in rural areas it is crucial to promoting accessibility and social inclusion (Berg & Ihlström, 2019). The transportation sector was the largest greenhouse gas emitting sector in the UK contributing 26% of the total emissions in 2021. Private car use is considered particularly problematic with estimates suggesting that a journey made with a typical petrol car generates four times more CO_2 per person than the same journey made by a coach (DfT, 2023). Not only do cars contribute significantly to greenhouse gas emissions they also compete for space, are responsible for congestion, noise, localised heat effects and in some cases fatal accidents (Gössling, 2020). Increasing PT ridership is considered key to a more sustainable future for transport (Spandou & Macário, 2021). Public receptiveness to using PT is directly related to their attitudes towards it (Soza-Parra et al., 2019). A key factor negatively affecting attitudes towards PT are disruption events and previous studies have demonstrated that short-term and long-term reductions in ridership follow disruption to PT networks (Nguyen-Phuoc et al., 2018; van Exel & Rietveld, 2009; Zhu et al., 2017). Life events also influence travel behaviour and having young children has been shown to increase car dependency (Prillwitz et al., 2006; Ryley, 2006). Census data showed that in 2023, 43% (8.3 million) of families in the UK had one or more dependent children (ONS, 2024). Travelling with young children represents a unique encumbrance (something that restricts or hinders freedom of movement) which can deter public transport use, for example, carrying child related paraphernalia (Price & Matthews, 2013), safety constraints, concerns around unpredictable behaviour and their effects on other passengers (McCarthy et al., 2017). Moreover, this group of passengers are particularly vulnerable to unexpected events such as disruption. The aim of this study was to understand the unique challenges faced by adults travelling with young children during a PT disruption event. Based on these findings a series of recommendations were developed, which will be of interest to policy makers and transport planners. The recommendations offer a tangible means to improve PT experience for this substantial group of potential PT passengers.

Approach

To better understand the challenges faced by adults encumbered by young children - a series of interviews were conducted with adults who have experience of travelling with young children. A hypothetical scenario was developed, based on the results of a previously conducted survey (unpublished data). Participants were informed that on arriving at a train station that their train journey had been delayed by 45 minutes. Participants were then asked what they would choose to do when travelling with a young child in a pushchair and how this would differ to when travelling alone. Semi structured questioning was developed which probed the factors which would affect responses and decision making. A total of 20 interviews were conducted via Microsoft Teams, the interviews lasted approximately 30 minutes and were auto transcribed. Thematic analysis as set out by Braun and Clarke (2021) was used to analyse the data and generate themes which captured the data findings. An inductive approach focussing on semantic themes was used to develop themes which summarised the original research question.

Results

Thematic analysis of participant interview transcripts led to the development of eight themes; comfort, flexibility, impact, information, journey planning and decision making, practical and logistical considerations, safety and emotional and social impact. Owing to space constraints the theme of flexibility and subthemes are shown in Table 1, along with relevant descriptions and a series of recommendations. This will be expanded to cover all themes for the conference output.

Table 1: Subthemes, descriptions and recommendations for flexibility theme.

Theme	Subtheme	Description	Recommendations
Flexibility	Exclusion	Passengers travelling with children conveyed a feeling of exclusion from alternative travel modes because of accessibility issues or unsuitability.	1. Car seat provision to allow safe use of taxis 2. Improved connectivity and accessibility between services for example shuttles to connect trains and buses and lifts/travelators to improve step free access. 3. Improved visual/audio information on alternatives and accessibility to enable hands free information gathering
	Urgency	Passengers travelling alone were often concerned with journey efficiency and used private taxis as a back-up. Passengers travelling with children conveyed that they would usually be travelling for leisure, with less time pressure.	
	Adaptable	Passengers travelling alone had agency to investigate other travel modes and flexibility in their mobility. Passengers travelling with children commented that inability to access relevant information or access alternative provision would force them to wait and stick to the original plan.	

Key findings

Adults travelling with young children generally made different choices to when travelling alone based on a myriad of factors, including reduced flexibility. The recommendations suggested here focus primarily on improving the experience of adults travelling with young children. However improved accessibility will benefit passengers with mobility issues and audio/visual information (that doesn't require access to mobile phone) will benefit a wide range of passengers. The full set of recommendations developed by this study, informed by all themes, therefore has the potential to make PT networks more inclusive and attractive to a wide range of passenger groups.

References

Berg, J., & Ihlström, J. (2019). The Importance of Public Transport for Mobility and Everyday Activities among Rural Residents. *Social Sciences*, *8*(2), 58. https://www.mdpi.com/2076-0760/8/2/58

Braun, V., & Clarke, V. (2021). *Thematic Analysis: A Practical Guide*. SAGE Publications. https://books.google.co.uk/books?id=mToqEAAAQBAJ

DfT. (2023). *Transport and environment statistics: 2023*. Department for Transport. Retrieved 25.04.24 from https://www.gov.uk/government/statistics/transport-and-environment-statistics-2023

Gössling, S. (2020). Why cities need to take road space from cars - and how this could be done. *Journal of Urban Design*, *25*(4), 443-448. https://doi.org/10.1080/13574809.2020.1727318

McCarthy, L., Delbosc, A., Currie, G., & Molloy, A. (2017). Factors influencing travel mode choice among families with young children (aged 0–4): a review of the literature. *Transport Reviews*, *37*(6), 767-781. https://doi.org/https://doi.org/10.1080/01441647.2017.1354942

Nguyen-Phuoc, D. Q., Currie, G., De Gruyter, C., & Young, W. (2018). Transit user reactions to major service withdrawal – A behavioural study. *Transport Policy*, *64*, 29-37. https://doi.org/https://doi.org/10.1016/j.tranpol.2018.01.004

ONS. (2024). *Families and househiolds in the UK: 2023*. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/bulletins/familiesandhouseholds/2023

Price, L., & Matthews, B. (2013). Travel time as quality time: parental attitudes to long distance travel with young children. *Journal of Transport Geography*, *32*, 49-55. https://doi.org/https://doi.org/10.1016/j.jtrangeo.2013.08.001

Prillwitz, J., Harms, S., & Lanzendorf, M. (2006). Impact of Life-Course Events on Car Ownership. *Transportation Research Record*, *1985*, 71-77. https://doi.org/10.3141/1985-08

Ryley, T. (2006). Use of non-motorised modes and life stage in Edinburgh. *Journal of Transport Geography - J TRANSP GEOGR*, *14*, 367-375. https://doi.org/10.1016/j.jtrangeo.2005.10.001

Soza-Parra, J., Raveau, S., Muñoz, J. C., & Cats, O. (2019). The underlying effect of public transport reliability on users' satisfaction. *Transportation Research Part A: Policy and Practice*, *126*, 83-93. https://doi.org/https://doi.org/10.1016/j.tra.2019.06.004

Spandou, M., & Macário, R. (2021). Public Transport's Contribution Toward Achieving the Sustainable Development Goals. In (pp. 847-865). https://doi.org/10.1007/978-3-319-95873-6_4

Sun, Y., & Cui, Y. (2018). Evaluating the coordinated development of economic, social and environmental benefits of urban public transportation infrastructure: Case study of four Chinese autonomous municipalities. *Transport Policy*, *66*, 116-126. https://doi.org/https://doi.org/10.1016/j.tranpol.2018.02.006

van Exel, N. J. A., & Rietveld, P. (2009). When strike comes to town…anticipated and actual behavioural reactions to a one-day, pre-announced, complete rail strike in the Netherlands. *Transportation Research Part A: Policy and Practice*, *43*(5), 526-535. https://doi.org/https://doi.org/10.1016/j.tra.2009.01.003

Zhu, S., Masud, H., Xiong, C., Yang, Z., Pan, Y., & Zhang, L. (2017). Travel Behavior Reactions to Transit Service Disruptions: Study of Metro SafeTrack Projects in Washington, D.C. *Transportation Research Record*, *2649*(1), 79-88. https://doi.org/10.3141/2649-09

Impact of Neck Support on Headrests for Enhancing Relaxation and Comfort

Mohsen Zare[1], Nahed Jaffel[2], Hugues Baume[1] & Fabien Bernard[1]

[1]UTBM, France, [2]University of Monastir, Tunisia

SUMMARY

This study explores the impact of neck support in headrest design to enhance relaxation and comfort, particularly in autonomous vehicles. A dual-phase approach was employed for iterative development and evaluation of headrest prototypes, aiming to identify ergonomic improvements and optimise comfort for diverse user profiles. In the preliminary phase, the comfort of headrests from Peugeot 3008, Citroën C4, and DS7 Crossback vehicles was evaluated. Three headrest models were tested using subjective questionnaires (CP50 scale and body map) and objective biomechanical measurements, including pressure mapping, electromyography (EMG), and video recording. Five volunteers participated, performing activities such as napping and reading, while neck and shoulder muscle activity was assessed. In the second phase, the standard headrest was compared with two newly designed prototypes using the same protocol. Results showed that headrests with enhanced neck support significantly reduced neck muscle activity and increased head-to-headrest contact compared to classic designs. These findings highlight the need for redesigned headrests to improve onboard comfort in next-generation vehicles.

KEYWORDS

Neck support, headrests, relaxation, comfort, autonomous vehicles

Introduction

Comfort is a critical factor in vehicle design, as consumers increasingly prioritise long-term usability when selecting a car. Manufacturers strive to deliver comfortable products to maintain a competitive edge (Vink, 2004; Zenk et al., 2011). Among the key elements of vehicle ergonomics, neck support provided by headrests plays a pivotal role in maintaining proper posture, reducing muscle strain, and preventing discomfort or fatigue during extended use.

Previous research has examined the relationship between headrest design and comfort, focusing on factors such as shape, foam density, and adjustability (Bouwens et al., 2018; Franz et al., 2012a). These studies underscore the importance of ergonomic design in enhancing relaxation and comfort during long drives. However, there remains no consensus on the optimal headrest design to accommodate diverse body types and driving conditions.

This study aims to address these gaps by examining how headrest designs influence muscle activity and perceived comfort across varied user profiles. A comprehensive assessment approach was adopted, combining EMG, pressure mapping, and subjective questionnaires. The combination of these methods provides a multidimensional evaluation framework for a holistic evaluation of the user comfort.

Methods

This study evaluated the ergonomic comfort of headrests in two phases. The first phase assessed three existing headrests from the Peugeot 3008, Citroën C4 Picasso, and DS7 Crossback vehicles. Objective measures included EMG for neck and upper trapezius muscle activity and pressure mapping to assess the distribution and magnitude of support provided by the headrests. Subjective evaluations were collected via the CP50 scale and a custom comfort survey. Five participants, representing diverse anthropometric profiles (5th to 95th percentiles), were recruited for this phase.

Based on the findings of the first phase, two new prototype headrests were developed with enhanced neck support features. In the second phase, these prototypes were tested alongside the standard Peugeot 3008 headrest. The same protocol was applied, involving 20-minute sessions of reading and napping on each headrest. Participants provided subjective feedback immediately after each activity, and objective data were collected to ensure consistency.

Three participants from the original cohort participated in the second phase, while two withdrew due to scheduling conflicts. Anthropometric measurements were recorded again for consistency. EMG sensors and pressure mapping techniques were used to capture data on neck and shoulder muscle activity and the distribution of support. Statistical analyses were performed to compare these metrics across all tested designs. Figure 1 presents the headrests from the Peugeot 3008, Citroën C4 Picasso, and DS7 Crossback used in this study, along with the newly developed prototypes.

Peugeot 3008 | DS 7 Crossback | Citroën C4 Picasso | Prototype 1 | Prototype 2

Figure 1: Headrests of the Peugeot 3008, Citroën C4 Picasso, DS7 Crossback and two newly developed prototypes

Results

The results indicate that headrests with adjustable features along the X-axis provided superior neck support compared to the standard models, which only offered Z-axis adjustability. Contact surface areas were significantly larger with the adjustable headrests, while the standard headrest exhibited higher pressure concentrations. All tested headrest models revealed limitations in providing sufficient neck contact and support.

EMG data suggested trends toward reduced neck muscle activity with the enhanced designs, though definitive conclusions were limited by the small sample size. The second phase demonstrated that the new prototypes significantly improved neck support and reduced muscle activity compared to the standard model.

Conclusion

This study highlights the necessity of designing headrests that prioritise both safety and comfort, particularly in the context of autonomous vehicles where prolonged seating is common. Headrests with enhanced adjustability provide larger contact surfaces and reduced muscle activity, while standard models exhibited higher pressure. The newly developed prototypes demonstrated improved comfort and ergonomic support, underscoring their potential for next-generation vehicle designs.

While this study primarily focused on comfort and relaxation in headrest design, it is important to acknowledge the dual role of headrests in both comfort and occupant safety. The enhanced neck support in our prototype designs may have implications for safety performance, particularly in reducing head movement during rear-impact crashes. However, the present study did not include crash testing, and future research should investigate the impact of these designs on occupant safety in crash scenarios.

Furthermore, the small sample size in this study specially in balanced gender subjects presents limitations for generalisation of results. However, participants were selected to represent a broad range (5th to 95th percentiles) to enhance generalisability. Future studies with a larger sample size should aim to further explore gender-based variations in perceived comfort and biomechanical responses.

References

Bouwens, J. M. A., Schultheis, U. W., Hiemstra-van Mastrigt, S., & Vink, P. (2018). Expected versus experienced neck comfort. Human Factors and Ergonomics in Manufacturing & Service Industries, 28(1), 29-37. https://doi.org/10.1002/hfm.20721

De Looze, M. P., Kuijt-Evers, L. F. M., & van Dieën, J. (2003). Sitting comfort and discomfort and the relationships with objective measures. Ergonomics, 46(10), 985-997. https://doi.org/10.1080/0014013031000121977

Franz, M., Durt, A., Zenk, R., & Desmet, P. M. A. (2012a). Comfort effects of a new car headrest with neck support. Applied Ergonomics, 43(2), 336-343. https://doi.org/10.1016/j.apergo.2011.06.009

Franz, M., Durt, A., Zenk, R., & Desmet, P. M. A. (2012b). Comfort effects of a new car headrest with neck support. Applied Ergonomics, 43(2), 336-343. https://doi.org/10.1016/j.apergo.2011.06.009

Liu, Y., Liu, Q., Lv, C., Zheng, M., & Ji, X. (2018). A Study on Objective Evaluation of Vehicle Steering Comfort Based on Driver's Electromyogram and Movement Trajectory. IEEE Transactions on Human-Machine Systems, 48(1), 41-49. https://doi.org/10.1109/THMS.2017.2755469

Trojaniello, D., Cristiano, A., Oleari, E., Tettamanti, A., & Sanna, A. (2018). Car seat comfort assessment based on objective and subjective measurements in elderly population.

Vink, P. (Éd.). (2004). Comfort and Design: Principles and Good Practice. CRC Press. https://doi.org/10.1201/9781420038132

Zenk, R., Franz, M., Bubb, H., & Vink, P. (2011). Technical note: Spine loading in automotive seating. Applied ergonomics, 43, 290-295. https://doi.org/10.1016/j.apergo.2011.06.004

Zhang, L., Helander, M. G., & Drury, C. G. (1996). Identifying factors of comfort and discomfort in sitting. Human Factors, 38(3), 377-389. https://doi.org/10.1518/001872096778701962.

Integration of personas in transport policymaking

Phuong Anh Nguyen, Robert Houghton, Amanda Crompton & Sarah Sharples

University of Nottingham, Nottingham, UK

SUMMARY

There is increasing interest in the concept of human-centred design (HCD) for policymaking and the application of Human Factors and Ergonomics (HF/E) methods in this area. In this study, we explored perceptions of a specific method, mainly personas, in the transport policy space. This research aim was accomplished through sixteen interviews with transport policymakers and analysts working in both national and local government, employing critical decision methods and thematic analysis to examine the transcripts.

KEYWORDS

Human-centred design, policymaking, transport, personas

Research background

Using personas is one human-centred design (HCD) technique (Friess, 2012), which has predominantly been applied in designing commercial products and services, while a small amount of research has demonstrated the potential of using personas for social good and policymaking (Gonzalez de Heredia et al., 2018). The application of HCD tools in policymaking, however, faces several challenges, including issues of stakeholder inclusion, alignment with existing policy systems, and the potential for oversimplification (Nguyen et al., 2024).

The research gap arises in both policy studies and HFE research. Current policy research is mainly concerned with the potential implication of design approaches in public policy rather than discussing the utilisation of a specific design technique. In HFE research, although extensive studies exist about the utility of data and methodology for creating personas and the advantages and the challenges of applying personas to design projects (Goh et al., 2017), there is a notable scarcity of research on how personas are involved in the decision-making process, accompanied by a deficiency of empirical evidence and variable practical impact (Chapman & Milham, 2006; Rönkkö et al., 2004). National and local governments have developed transport user personas to comprehend commuters' behaviour. Integrating this tool and other potential HCD techniques in policymaking is a critical problem in policy studies and HFE, an under-studied research topic.

To integrate HCD techniques into well-established, structured and sophisticated policymaking, it is vital to understand the perspective of the personnel who will create and use those tools to deal with actual policy issues. This research proposes to fill the gaps in comprehending how personas could be used to bring attention to human values in transport policymaking. The objective is to investigate policymakers' receptiveness to using personas, identify appropriate policy contexts for their use, and then give insight into how to develop personas to improve their usability in policymaking.

Research methods

A diverse group of policymakers and analysts who directly work with policymakers in both national and local government, including the Department for Transport, National Highway, Transport for London, Transport for Manchester, and East Midlands Combined Country Authority, possessing diverse roles and varying levels of expertise, were invited to join the interview. The duration of each interview ranges from 30 minutes to 1 hour. The interviews are structured according to the principles of the Critical Decision Method (CDM), which is a cognitive task analysis technique developed by Klein et al. (1989). The participants were asked to share about a project that they have done, followed by discussion questions on HCD to capture the knowledge, strategies used, and decision-making process in complex, real-world situations.

There were sixteen interviews conducted online via the Teams platform and then transcribed for analysis. To analyse the qualitative data, the thematic analysis is conducted in six phases presented by Clarke & Braun (2017).

Brief research findings and discussion

The interview result shows that HCD is a novel concept for all interviewed participants, while they are more familiar with user-centred design and highlight the influence of policy on citizens. However, when discussing the policymaking process, the participant pointed out its complexity, emphasising the significance of engaging with internal teams and multiple transportation stakeholders. Policymaking is a structured yet flexible process that necessitates a broader context, thorough data, and the interconnectedness of numerous aspects in analysis, which significantly depend on personal experience and intuition. Moreover, senior policymakers, as leaders, consider the well-being and workload of their teams. In a policy team, it is critical to incorporate a variety of experiences and profiles while explicitly acknowledging the value of various working styles, as this fosters the greatest diversity and enhances the team's effectiveness. Furthermore, in certain policy projects, additional social aspects, including animals, are also considered. Consequently, policymaking inevitably necessitates an HCD approach and an empathic thinking process.

Regarding the utilisation of personas, opinions are varied, while some policymakers stand strongly for using personas in policymaking. While others express concerns about potential oversimplification. The existing transport user personas predominantly focus on surface transport users, which is less pertinent to other projects, particularly if those projects do not entail direct public interaction. Thus, the tools need to provide appropriate evidence for decision-making. Additionally, integrating personas into the policymaking process presents significant challenges, which necessitates clear leadership expectations and compelling case studies to illustrate their efficacy. It is essential to integrate personas into daily practices, transforming them into a regular tool rather than an occasional one.

Personas have great potential to put humans at the heart of policymaking. However, to achieve this goal, the creation and usage of personas should meet multiple requirements. First, the data used to develop personas must be obtained accurately from multiple sources that reflect genuine experiences and needs of transport users of different groups, especially marginally or under-presented groups. It is essential to contemplate the potential evolution of demographics and wants over the next 10-20 years and ensure that personas are subject to updates and adaptations when new facts and insights emerge. Moreover, training and guidance are important to ensure that all team members comprehend the efficient use of personas and are willing to use them in the teamwork process. Furthermore, given that policymaking is both structured and flexible, the presentation of personas must avoid oversimplification and provide appropriate evidence while stimulating open-mindedness for innovative ideas and promoting empathy, inclusivity, and strategic clarity in policy design, ultimately cultivating more equitable and effective policy outcomes.

By exploring how to integrate personas in policymaking, the research results contribute to enhancing policymakers' work, hence improving the policies which reflect the demands of diverse communities. This will help extend the impact of HF/E for wider societal advantages.

References

Chapman, C. N., & Milham, R. P. (2006). The Personas' New Clothes: Methodological and Practical Arguments against a Popular Method. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50(5), 634–636. https://doi.org/10.1177/154193120605000503

Clarke, V., & Braun, V. (2017). Thematic analysis in psychology. The Journal of Positive Psychology, 12(3), 297–298. https://doi.org/10.1080/17439760.2016.1262613

Friess, E. (2012). Personas and decision making in the design process. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1209–1218. https://doi.org/10.1145/2207676.2208572

Goh, C. H., Kulathuramaiyer, N., & Zaman, T. (2017). Riding Waves of Change: A Review of Personas Research Landscape Based on the Three Waves of HCI (pp. 605–616). https://doi.org/10.1007/978-3-319-59111-7_49

Gonzalez de Heredia, A., Goodman-Deane, J., Waller, S., Clarkson, P. J., Justel, D., Iriarte, I., & Hernández, J. (2018). Personas for policy-making and healthcare design. 2645–2656. https://doi.org/10.21278/idc.2018.0438

Klein, G. A., Calderwood, R., & MacGregor, D. (1989). Critical decision method for eliciting knowledge. IEEE Transactions on Systems, Man, and Cybernetics, 19(3), 462–472. https://doi.org/10.1109/21.31053

Nguyen, P.-A., Houghton, R., Crompton, A., & Sharples, S. (2024). WIP - Human-centred design in Policymaking: Bibliometric analysis and systematic review. Proceedings of the European Conference on Cognitive Ergonomics 2024, 1–7. https://doi.org/10.1145/3673805.3673852

Rönkkö, K., Hellman, M., Kilander, B., & Dittrich, Y. (2004). Personas is not applicable: local remedies interpreted in a wider context. Proceedings of the Eighth Conference on Participatory Design Artful Integration: Interweaving Media, Materials and Practices - PDC 04, 112. https://doi.org/10.1145/1011870.1011884

Project PRIME: Enhancing human performance with motorcycle road markings on regular roads

Alex Stedmon[1], David McKenzie[2], Martin Langham[1], Kevin McKechnie[3], Richard Perry[2], Stuart Wlison[2], Morag Mackay[2], Stuart Geddes[3] & Vince Tait[2]

[1]Open Road Simulation Ltd, UK [2]Transport Scotland, UK [3]BEAR Scotland, UK

SUMMARY

Project PRIME is investigating new road markings that demonstrate positive behaviour change for motorcycle casualty reduction. These road markings have been designed as a low-cost solution for road safety and a key research question is whether they are suitable for regular roads that have not been improved through expensive engineering works. This paper reports the findings for regular roads where these unique road markings produced statistically significant reductions in motorcycle speed, improved lateral lane position, and reduced braking and increased use of the markings across collision cluster sites on key motorcycle routes across the West Highlands of Scotland. These findings provide confidence that PRIMEs can be installed on regular roads where motorcycle collisions need to be addressed. Project PRIME has demonstrated major impact in rider safety and was awarded the Prince Michael International Road Safety Award in 2023.

KEYWORDS

Transport human factors, Motorcycle rider, Behaviour change, Road markings, Casualty reduction

Introduction

In the UK, between 2015 and 2020, an average of six motorcyclists were killed and 115 were seriously injured each week in reported road casualties (Department for Transport 2022). This is a picture that is repeated around the world with motorcyclists grossly over-represented in road traffic collision statistics (de Moraes, Godin, Dos Reis, Belloti and Bhandari, 2014; Vanlaar, Hing, Brown, McAteer, Crain and McFaull, 2016, Transport Scotland, 2021).

Motorcyclists are notoriously difficult to engage with road safety initiatives and riders who have not taken further training may develop poor riding styles and lack riding skills to keep them safe. In many incidents, the motorcyclist is the sole vehicle involved and common causes are attributed to a poor turn or manoeuvre, exceeding the speed limit, loss of control, travelling too fast for the conditions or sudden braking (Department for Transport 2021).

In an attempt to reduce road casualties, many governments, road safety agencies and practitioners around the world are adopting the Safe System approach which promotes a transport system free from death and serious injury (Job, Truong, and Sakashita, 2022; Transport Scotland, 2021). A key focus is the fallibility of humans so that the transport system is more supportive of their needs and wider policy requirements.

There are five guiding principles to the Safe System: people make mistakes; people are vulnerable to injury; shared responsibility for road safety; forgiving systems approach; and vision zero where the aim is for no one is killed or seriously injured on the roads (Job, Truong, and Sakashita, 2022).

From these principles, five inter-related pillars have been developed for Safe System design: safe road use, safe roads and roadsides, safe speeds, safe vehicles, better post-crash responses. The current research was conducted in accordance with the Safe System approach and the findings are considered in respect of this in the discussion.

PRIME road markings to enhance human performance

Dedicated road markings, designed as 'Perceptual Counter-Measures' (PCMs) have been shown to influence road user behaviour. These are typically road markings that dictate a desired behaviour by altering how a road user might perceive and process risk factors in the environment around them (Gardener, Tate, Mackie, Stedmon, and Southey-Jones, 2017; Mulvihill, Candappa, and Corben, 2008).

With the current research, a new approach was taken by developing a tool for motorcyclists through the design of 'Perceptual Rider Information to Maximise Expertise and Enjoyment' (PRIMEs). The underlying philosophy of PRIMEs is to provide a road behaviour solution for motorcycle riders using terms that are relevant to them such as expertise and enjoyment (Stedmon, McKenzie, Langham, McKechnie, Perry and Wilson, 2021, 2022). It is also important that as a casualty reduction intervention, PRIMEs are cost effective to install and maintain. PRIMEs should be able to be installed on existing (i.e. regular) roads or incorporated into larger road upgrade schemes.

For motorcyclists safely navigating bends it is important that: speed is suitable for the conditions, position is optimised for entering and travelling through the bend, and braking is minimised. A combined road sign and road marking were investigated in this research (Figure 1).

Figure 1: PRIMEs road sign and 'gateway' road marking

The road marking was designed as a series of three 'gateway' markings positioned on the approach to a bend. The intention was that the road marking would encourage motorcyclists to ride 'through the gap' and use the gateways as a cue to adjust their riding behaviour. Throughout the research programme, the results have demonstrated positive behaviour change (Stedmon, et al, 2021, 2022; Stedmon, McKenzie, Langham, McKechnie, Perry, Wilson, Mackay and Geddes, 2024).

With international interest from road authorities wishing to install PRIME road markings and a keen awareness of limited budgets for road safety, a key question for the research was whether PRIME road markings would be effective on regular roads. In Phase 1 of Project PRIME the trial sites were brought up to the best standard through engineering works to support the research. Moving on from this, trial sites were identified where no works were conducted and which represented typical regular roads where other authorities might install PRIME road markings.

Trial site selection

An initial analysis by Transport Scotland of the Trunk Road Network identified 660 collisions involving motorcyclists between 2013 and 2017. Using STATS19 data (reported directly from Police attending accident scenes) the North-West region was identified as a priority area for motorcycle casualty reduction. BEAR Scotland Ltd (North-West Unit) conducted a further review of collision cluster sites. Between 2008 and 2017, sites within a 100m radius where three or more personal injury accidents (PIAs) involving a motorcyclist or pillion highlighted the A82, A85 and A83 as priority routes (BEAR Scotland, 2019).

Six trial sites were selected for these road trials and were spread over a large geographic area ranging from Crianlarich, Crieff and Stirling. The site represented a range of bends on rural roads with speed limits over 40mph in line with recent casualty statistics (Transport Scotland, 2020) and a range of characteristics (e.g. complex geometry, tightening apexes, descents and inclines prior to bends, bends off fast sections of road). The sites are represented below (Figure 2).

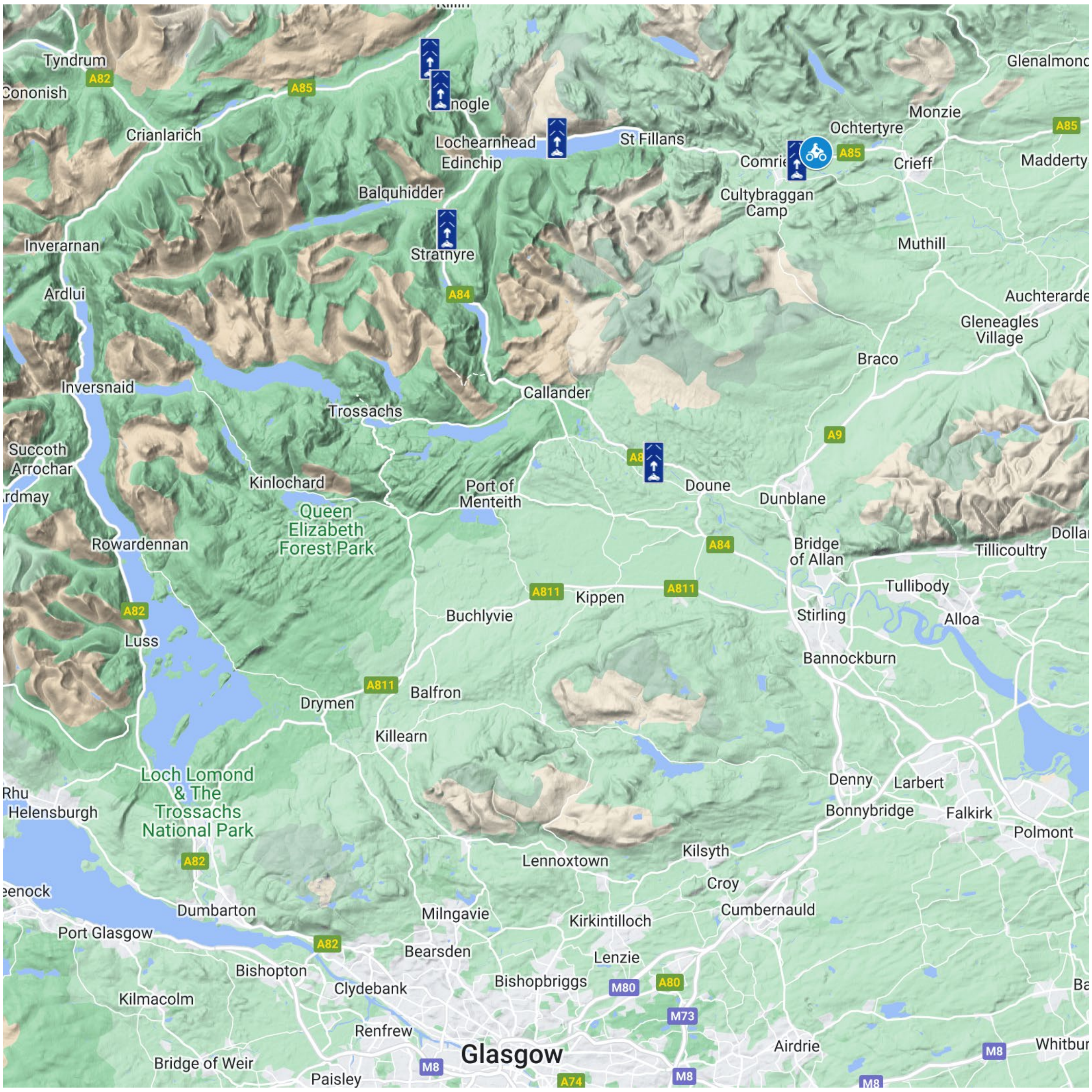


Figure 2: PRIMEs trial sites

A comparison site was also included where data were collected but the PRIME road markings were not installed. Due to the wide variety of bends and road characteristics on the Trunk Road Network, the comparison site was not regarded as experimental control condition (i.e. where identical conditions are usually compared statistically).

Method

This research followed a conventional 'pre- and post-intervention' quasi-experimental approach. Baseline data were compared with data collected once the PRIME road markings had been installed.

Participants

This research relied on an opportunistic sample of motorcyclists. Across the trial sites 4,652 motorcycles were observed and from these 1,542 lead motorcycles were analysed in more detail.

Apparatus

Data were captured at each site using small and inconspicuous weatherproof video cameras typically attached to roadside posts or trees. The cameras captured 1080p video at 60Hz for time periods of at least 20hrs, stored on 512Gb microSD cards. At each site, three cameras were installed facing: towards the rider, behind the rider and perpendicular to the rider a short distance ahead of the last PRIME road marking.

The PRIME road markings were installed using 3M™ Stamark™ High Performance permanent tape. This material provided increased visibility, grip and safety, even in the wet. It also offered high levels of adhesion to the road surface and provided a permanent marking that would not be disturbed by other vehicles (i.e. general traffic and heavy goods vehicles).

Design

The independent variable in this research was the PRIME road markings which had two levels: Baseline (without PRIMEs installed) and PRIME (with PRIMEs installed).

Baseline and PRIME data were collected on a number of occasions, as specified below:

- Baseline 1 and 2 – two separate weekends before PRIMEs were installed
- PRIME 1 – the weekend after PRIMEs were initially installed
- PRIME 2 – six or eight weeks after the PRIME 1 data collection

A range of dependent variables were identified to capture data about the potential influence of PRIMEs on rider behaviour, including speed, lateral position, braking and use of the final PRIME road marking.

Procedure

In a change from the previous research, prior to data collection, trial sites were not upgraded (i.e. resurfacing and new vehicle restraint systems). This was important in order to evaluate PRIMEs on regular roads. Care was also taken to make sure that no changes to the sites were undertaken during the trials (i.e. scheduled road works).

Data were captured during the typical motorcycle season (i.e. May to September) when motorcyclists were most active. Weekends were chosen for data collection as this was generally when motorcyclists ride for leisure/social purposes. During each weekend cameras were set up at the trial sites and recorded all road traffic during Saturday and Sunday from 09:00 to 17:00. Power supplies were replenished through the weekend and cameras were collected on Sunday evenings.

Ethics and risk assessment

An independent review of potential ethical issues was conducted. Approval was granted in accordance with general principles of the British Psychological Society and International protocols. A risk assessment was also conducted in order to safeguard the research activities. Induction training was undertaken so that roadside safety protocols were adhered to and the correct PPE was worn at all times. The design for the PRIME road markings and road sign went through a formal application process for authorisation of non-prescribed traffic signs (Road Traffic Regulations Act 1984: Sections 64 and 65). Approval was granted prior to the trials taking place. Following on from this, independent road safety audits are conducted at regular intervals to oversee the safe installation of PRIMEs at all trial sites.

Data analyses

Baseline 1 and Baseline 2 datasets were compared by conducting a T-Test (t) to identify any differences between them. Where any significant differences were observed, effect size was calculated using Cohen's (ds) equation. Where the Baseline 1 and Baseline 2 datasets were observed to be the same (i.e. there was no significant difference) they were combined into a single dataset (i.e. 'Baseline'). This was the case in all baseline comparisons reported in this paper.

Speed and lateral position data were analysed using one-way Analysis of Variance (ANOVA) techniques. Where any significant results were observed, effect size was calculated using a partial eta squared (η^2) analysis. Post-hoc Bonferonni-Hoch analyses were conducted in order to determine where significant differences occurred between the datasets. Tests for effect size were conducted using Cohen's (ds) calculations.

Braking behaviour and use of PRIMEs datasets were analysed using Chi Square (X^2) tests. Where any significant results were observed, effect size was analysed using Cramér's V (V) calculations. Further post-hoc analyses were performed by calculating standardised residuals in order to determine where significant differences occurred between the datasets.

In previous analyses intra-rater reliability was assessed instead of inter-rater reliability (Stedmon et al, 2021). This followed the process set out by Mackey and Gass (2005) where ratings were conducted at different time intervals (i.e. T1 and T2) and then analysed in the same way as inter-rater reliability. Cohen's Kappa (k) calculations were conducted for samples of data for speed, lateral position and braking in the 2020 road trials. For speed and braking perfect matches were observed (k=1.0) due to the discrete nature of these data. For lateral position k=0.92 indicating a very high agreement and only minor differences in coding at the thresholds between the three lane positions.

Results

In total 4,652 motorcycles were processed across all the trial sites and from these 1,542 lead motorcycles were analysed in more detail. Results from the trial sites are summarised below (Table 1).

Table 1: Motorcycle numbers throughout the road trials

Trial site	Rider Behaviour				
	Speed	**Position at PRIME**	**Position at Apex**	**Braking**	**Use of Gateway**
Landrick	Sig	-	Sig	-	Sig
Strathyre	Sig	Sig	-	-	Sig
Glenogle	Sig	Sig/Trend	Sig	Sig/Trend	Sig
Mid Lix	Sig	Sig	Sig	-	Sig
Dalkenneth	Trend	Trend	Sig	Trend	Sig
West Lodge	Sig	Sig	Sig	Sig/Trend	Sig
Lawers Lodge (comparison)	-	-	-	-	-

These results are summarised below:

- **Speed** – statistically significant reductions in speed were observed at 5 of the 6 trial sites and a trend observed at the other trial site
- **Lateral position at the final PRIME road marking** – statistically significant changes in lateral position were observed at 4 of the 6 trial sites indicating that motorcyclists were riding in better positions on approach to the bend. Trends were observed at two trial sites
- **Lateral position at the apex of the bend** – statistically significant changes in lateral position were observed at 5 of the 6 trial sites
- **Braking behaviour** – statistically significant reductions in braking were observed at 2 of the 6 trial sites. Trends were observed at 3 trial sites
- **Use of the PRIME road markings** – statistically significant increases in the use of the road markings were observed at all 6 of the trial sites
- **Comparison site** – as expected, no differences in rider behaviour were observed.

A number of trends in the data were observed. A reduction in speed was found at Dalkenneth ($p=0.19$). For lateral position at the final road marking, at the Glenogle trial site the PRIME 1 result fell just outside of being statistically significant when the Bonferonni-Hoch corrections were applied (i.e. $p=0.026$ rather than $p<0.025$). A positive change in road position was also found at Dalkenneth ($p=0.11$). For position at the apex of the bend, at the Landrick trial site the PRIME 1 result fell just outside of being statistically significant when the Bonferonni-Hoch corrections were applied (i.e. $p=0.033$ rather than $p<0.025$). For braking, a reduction in braking on the bend was found at Glenogle ($p=0.15$) and Dalkenneth ($p=0.10$) and also for total braking at West Lodge ($p=0.13$).

Discussion

Overall, the results for the PRIME road trials provide evidence for a range of beneficial effects of PRIMEs on rider behaviour on regular roads. Across all key measures (i.e. speed, position, braking and use of the gateway) significant effects were observed at different sites during the trials.

There were no instances of statistically significant increases in speed, dangerous positioning, increases in braking or decreased use of the gateways. These observations provide further evidence that PRIMEs did not have a detrimental effect on rider behaviour. As such, even at locations where no statistically significant effects were observed, PRIMEs were no worse than not installing them at all.

Transport Scotland recently published its ‘Road Safety Framework to 2030′ (Transport Scotland, 2021). It proposes a Safe System approach to road safety delivery as set out in the National Transport Strategy Delivery Plan (Transport Scotland, 2020). In relation to the concept of PRIMEs, the current research addresses the following pillars:

- safe speeds – speed limits in a Safe System are designed for crash-avoidance and reducing physical impact. Key factors that should be taken into account in any decisions on local speed limits are history of collisions, road geometry and engineering, road function; composition of road users (including existing and potential levels of vulnerable road users); existing traffic speeds, and road environment (Transport Scotland, 2021). With these factors in mind, PRIMEs offer a potential tool for supporting speed limits on regular roads. With the observed reductions in speed and no statistically significant increases in speed, PRIMEs may therefore provide a means for maintaining safe speed limits rather than drastically reducing them. However, coupled with improved position on the road and reduced braking on bends this would appear to be supporting the rider experience more holistically rather than focusing on one specific measure of performance for safety.
- safe road use – road users should pay attention to the road ahead and the task in hand; adapting to the conditions (weather, the presence of other users, etc.); travel at lower speeds; and give sufficient room to all other road users, no matter what their mode of travel (Transport Scotland, 2021). PRIMEs appear to provide motorcyclists with a tool that allows them to adapt their behaviour to the road environment and which other road users may also use as a cue for demanding bends and the presence of motorcyclists. In this way PRIMEs may help ensure that road users are risk-aware, can develop coping strategies for demanding situations, and act appropriately to keep themselves and others safe on the road (Transport Scotland, 2021). This was demonstrated by the positive results for road position both at the final PRIME road marking and at the apex of the bend.
- safe roads and roadsides – the environment is designed to reduce the risk of collision and to mitigate the severity of injury should a collision occur. This can be achieved through design, maintenance and the implementation of strategies to reduce casualties on the roads (Transport Scotland, 2021). This can also be promoted through positive behaviours and safer sharing of spaces, the appropriate use of speed limits and signage that provides a much more affordable and sustainable way to protect the most vulnerable road users. PRIMEs provide a low-cost and easily maintained casualty reduction initiative working in harmony with other interventions such as bike-guard and other vehicle restraint system (VRS) solutions. They can be installed on existing roads quickly and efficiently or incorporated into road upgrade schemes. From the low incidence of braking across the trial sites, this would seem indicate that motorcyclists are generally set up well for these bends but that other effects on position and speed enhance safety further.

Across these strategic pillars PRIMEs have the potential to provide a new and unique contribution to the Safe System approach. There is clear evidence from the research that PRIMEs influence rider behaviour and alongside this an installation toolkit has been developed for others to use in their areas and support the roll-out of PRIMEs more widely.

Conclusion

This research has demonstrated that the PRIME and road markings produced behaviour change on regular roads. There were a range of positive effects through speed reduction, better road position, reduced braking and an increased use of the markings compared to the baseline. Furthermore, no negative behaviours arose from the installation of the road markings. This research expands the scientific knowledge base of previous findings (Stedmon et al, 2021, 2022, 2024) to support the wider incorporation of the new road markings in motorcycle casualty reduction within the Safe System approach to road safety.

References

BEAR Scotland Ltd. (2019). *North West Unit Motorcycle Accident Investigation*. (unpublished report prepared for Transport Scotland, Trunk Road and Bust Operations (TBRO) Directorate. Scheme Ref: 18/NW/0801/042

de Moraes, V. Y., Godin, K., Dos Reis, F. B., Belloti, J. C., and Bhandari, M. (2014). Status of road safety and injury burden: Brazil. *Journal of Orthopaedic Trauma* 28, S45-S46

Department for Transport (2021). *Reported Road Casualties in Great Britain: Motorcycle fact sheet, 2020*. The Stationery Office, London. https://www.gov.uk/government/statistics/reported-road-casualties-great-britain-motorcyclist-factsheet-2020/reported-road-casualties-in-great-britain-motorcycle-factsheet-2020

Department for Transport (2022). *Vehicle Speed Compliance Statistics for Great Britain 2021*. The Stationery Office, London. https://www.gov.uk/government/statistics/vehicle-speed-compliance-statistics-for-great-britain-2021/vehicle-speed-compliance-statistics-for-great-britain-2021

Gardener, R., Tate, F., Mackie, H., Stedmon, A. W., and Southey-Jones, B. (2017). Motorcycle-friendly roads: Applying a customer lens on the journey from identification to implementation. In, *Proceedings of Australasian Road Safety Conference 2017* 10-12 October 2017, Perth Australia.

Job, R. F. S., Truong, J., and Sakshita, C. (2022). The ultimate Safe System: redefining the Safe System approach for road safety. *Sustainability*, 14, 2978. doi.org/10.3390/su14052978

Mackey, A., and Gass, S.M. (2005). *Second Language Research: Methodology and Design*. Mahwah, NJ: Lawrence Erlbaum Associates.

Mulvihill, C., Candappa, N., and Corben, B. (2008). *Evaluation of Perceptual Countermeasures for Motorcyclists Stages 1 and 2: Final Report*. Victoria Australia: Monash University Accident Research Centre.

Stedmon, A. W., McKenzie, D., Langham, M., McKechnie, K., Perry, R., and Wilson, S. (2021). Safeguarding motorcyclists: Trialing new PRIME road markings for casualty reduction. *Transportation Research Part F: Psychology and Behaviour* 83(2021), 333-350

Stedmon, A. W., McKenzie, D., Langham, M., McKechnie, K., Perry, R., and Wilson, S. (2022). Safer road users: Investigating the influence of PRIME road markings on motorcycle rider behaviour. *Transportation Research Part F: Psychology and Behaviour* 83(2021), 368-385

Stedmon, A. W., McKenzie, D., Langham, M., McKechnie, K., Perry, R., Wilson, S., Mackay, M., and Geddes, S. (2024). Project PRIME: road markings for motorcycle casualty reduction (an overview of findings from 2020 to 2022). *Ergonomics* 1-16, https://doi.org/10.1080/00140139.2024.2361304

Transport Scotland (2020). *Reported road casualties Scotland 2019: Key findings report*. Transport Scotland, Glasgow: Crown Copyright. ISBN: 978-1-911672-04-3

Transport Scotland (2021). *Scotland's Road Safety Framework to 2030: Together making Scotland's roads safer*. Transport Scotland, Glasgow: Crown Copyright. ISBN: 978-1-911672-06-7

Vanlaar, W., Hing, M. M., Brown, S., McAteer, H., Crain, J., and McFaull, S. (2016). Fatal and serious injuries related to vulnerable road users in Canada. *Journal of Safety Research* 58, 67-77

Systems thinking for sustainability: Using ActorMaps to compare transport schemes

Rich C. McIlroy & John M. Preston

Transportation Research Group, University of Southampton

SUMMARY

Mobility as a Service (MaaS) is a smart mobility idea that aims to help people drive less by facilitating the combination of multiple modes of transport, planned, booked, paid for, and navigated through a single app. This apparent simplicity belies the complex structure of the underlying sociotechnical system. Here we take a systems ergonomics perspective, using Rasmussen's Risk Management Framework to analyse three current MaaS systems, i.e., Jelbi in Berlin, Floya in Brussels, and Breeze in the UK's Solent region, to proactively identify risks to success.

KEYWORDS

Mobility as a Service; Risk Management Framework; sociotechnical systems; sustainability

Introduction

Mobility as a Service (MaaS), a smart mobility system that brings together multimodal journey planning, ticketing, and route guidance, is an emerging, sustainable transportation concept. Current implementations provide a platform to plan multi-modal journeys (i.e., those involving two or more different transport modes) and buy tickets on a journey-by-journey basis. This research focusses on three real-world examples of such systems, applying a sociotechnical systems lens to compare the implementation of MaaS systems in Berlin (Germany), Brussels (Belgium), and the Solent region (south-central England) and map the stakeholders involved. In doing so, it offers insight into the similarities and differences between the implementations and provides a discussion of the organisational and governance structures we are likely to see dominate the future of Mobility as a Service. It does so in the theoretical context of Jens Rasmussen's Risk Management Framework (RMF; Rasmussen, 1997) with the aim of proactively identifying risks to the success of such schemes and, in turn, risks to transport sustainability. It takes a high-level, systems view and is based on discussions with key stakeholders in the three locations of interest.

AcciMaps and sustainability

The ActorMap tool, one part of the AcciMap method that is based on the RMF, was introduced by Rasmussen and Svedung (2000) as a way to visually represent the decision makers involved in a sociotechnical system and their hierarchical position in that system. It has previously been used as a stakeholder identification framework in the context of road safety (McIlroy et al., 2019); however, collisions are not the only negative externality of our transport system. There is merit in exploring the potential for the use of ActorMaps for stakeholder identification and framing in the context of sustainability. Therefore, just as Paul Salmon and colleagues did with their discussion of CWA and global challenges at the 2019 EHF conference (Salmon et al., 2019), we explore the use of a sociotechnical systems method in the context of sustainability.

The application of human factors and ergonomics (HFE) methods to sustainability challenges is by no means new, with Neville Moray calling for effort in this area over 30 years ago in his address to the International Ergonomics Association (Moray, 1995). Nevertheless, systems-based approaches to failure analysis are rarely applied to sustainability, despite this representing perhaps the greatest challenge (and most serious failure) faced by society today. The failure to achieve sustainability goals (in this context, the development and operation of a sustainable transport system) is one that systems ergonomics methods could help us understand and avoid.

With this impetus, this research uses the ActorMap tool as a form of proactive risk identification approach, in keeping with Rasmussen and Svedung's early descriptions of the approach (Rasmussen & Svedung, 2000). To this end, it uses the hierarchical system representation to facilitate the identification of actor and organisational related risks to the successful implementation of MaaS systems. Just as McIlroy et al. (2019) did in the context of road safety, we use the ActorMap in a comparative way, comparing the three MaaS implementations described above, i.e., the Breeze, Floya, and Jelbi systems of the Solent region, Brussels, and Berlin.

Method

To develop an ActorMap for each of the MaaS systems under analysis information was first gathered from publicly available transport and governmental websites. To validate the models, and to learn more about the functioning of each MaaS system, separate interviews were held with the programme managers for MaaS at each transport organisation running the scheme, i.e., BVG in Germany, STIB-MIVB in Belgium, and Solent Transport in the UK. In advance of each interview the initial ActorMaps were sent to each interviewee alongside an explanation of the method and the purpose of the study). Each interview lasted approximately one hour. Ethical approval for the study was sought from and granted by the University of Southampton's Ethical Committee (ID 93513). ActorMaps were edited based on notes made during the interviews, on the recordings, and on emailed responses to follow-up questions that were sent via email after the interviews were complete.

Results and discussion

The three completed actor maps are presented in Figures 1 to 3, below. Identified actors were categorised as governmental, public transport industry, other industry (including technology companies as well as non-transport focussed industry organisations), micromobility (including shared or hireable e-scooters and bikes), private hire transport (including car sharing or hiring and taxis), and end user and equipment. These categorisations are indicated using coloured shading in the figures below. The key is presented in Figure 4. The node with the thicker dashed surrounding line in each diagram indicates the main MaaS managing organisation in each setting.

Although laws, policies, standards, and guidelines set by organisations at the highest levels of the system are arguably most influential in terms of long-lasting system performance (i.e., by setting the conditions within which all else operates), the three interviewees all stated that there was minimal to no direct input from organisations at the highest national levels. The Breeze system in the UK differed in this regard only through its funding mechanism, with resources supporting the project coming directly from the Department for Transport. No other national governmental departments were identified as having a strong influence, though the interviewees did agree that the context of operation is shaped by policies set out by the organisations presented in the upper two levels, with funding decision taken at higher levels (e.g., on company car policies and public transport incentives) influencing use of the MaaS systems.

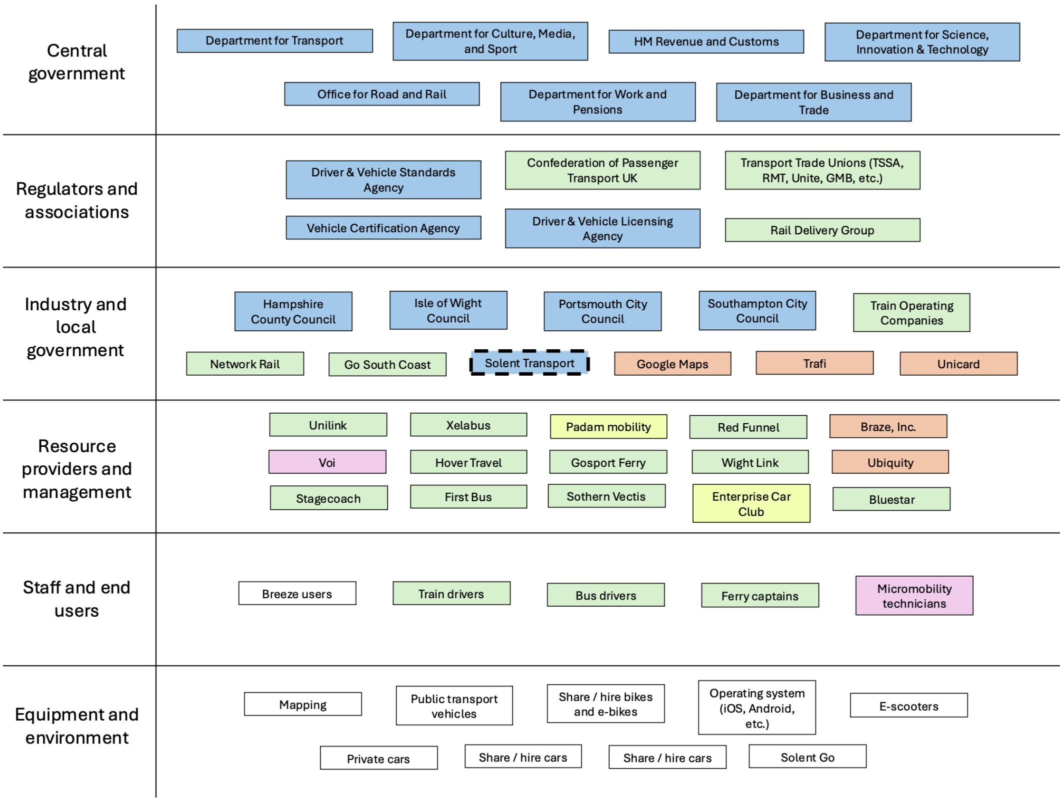


Figure 1. Actor map of the Solent region's 'Breeze' MaaS scheme. See Figure 4 for colour coding key.

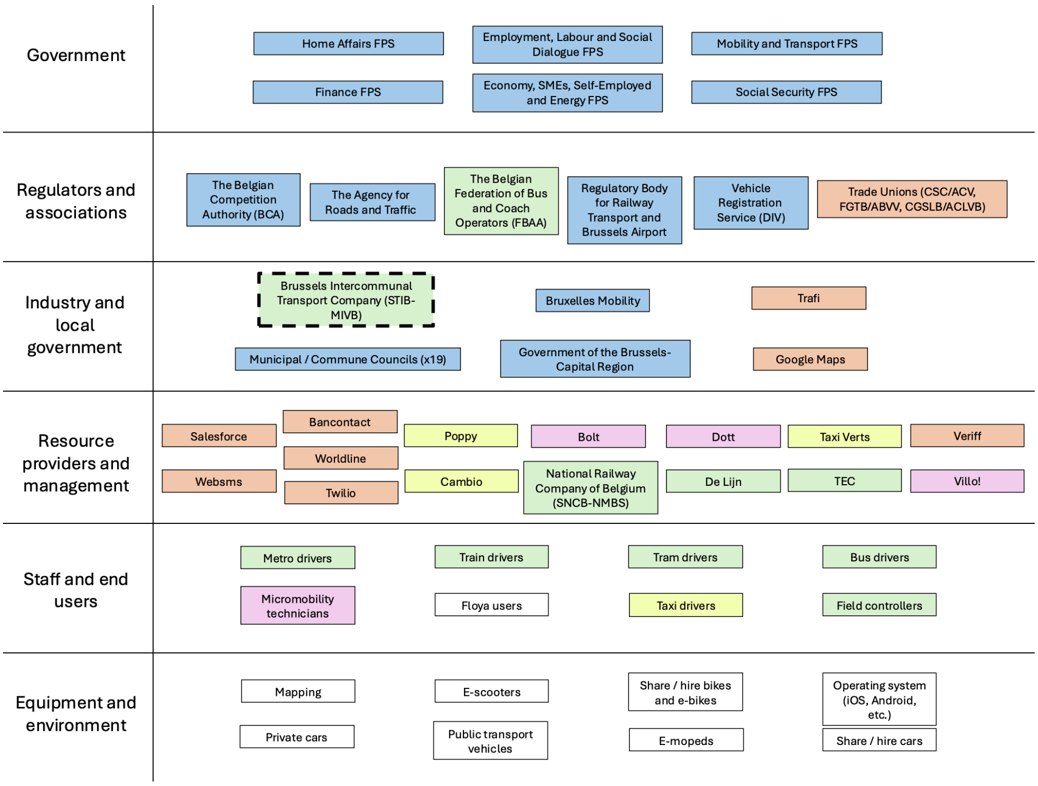


Figure 2. Actor map of the Brussels region's Floya MaaS scheme. See Figure 4 for colour coding key.

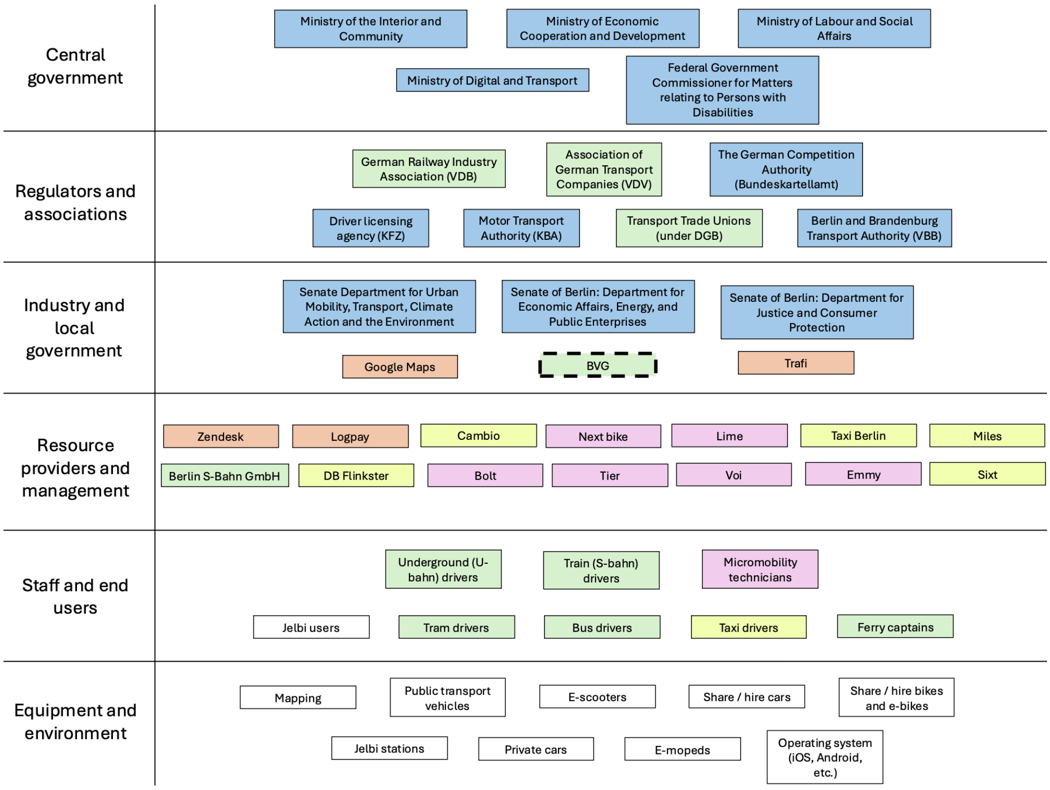


Figure 2. Actor map of the Berlin region's Jelbi MaaS scheme. See Figure 4 for colour coding key.

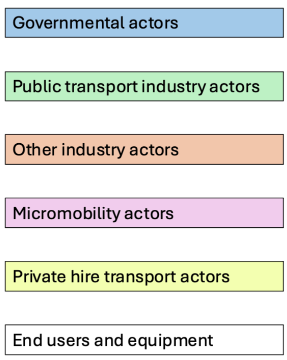


Figure 4. Actor Map colour coding scheme.

Influence from regulators and associations was noted by the Jelbi and Breeze programme managers in terms of the impact upon fare setting and regulation, with the Berlin and Brandenburg Transport Authority influential in this regard in Berlin, and the Rail Delivery Group having strict requirements for rail ticketing in the UK. The Association of German Transport Companies was also highlighted as placing requirements on digital ticketing and data exchange. In contrast, actors at this level were said to have only a very indirect influence over Floya in Brussels.

Beyond those examples, the three interviewees focussed primarily on actors residing at the central two levels, namely 'Industry and local government' and 'Resource providers and management'. These were the most immediately influential in terms of day-to-day system operations and were the levels where specific MaaS implementation differences were most notable (in addition to broader international governance and bureaucratic differences).

Although both Floya and Jelbi have more complex micromobility landscapes, with multiple companies operating and involved in the MaaS implementations, Breeze stood out as the most complex in terms of the governmental organisations and the number of public transport organisations involved. Public transport acts as the 'backbone' of MaaS (Alyavina et al., 2020; Mulley et al., 2023). The proliferation of separate companies therefore represents a major source of complexity and potential conflict, hence risk. In Brussels and Berlin, there is one main public transport company that runs local services, greatly simplifying the landscape in this regard. Moreover, and as mentioned above, those respective companies (i.e., STIM-MIVB and BVG) are the managers and operators of the MaaS systems in their jurisdictions. This is not the case in the Solent region where Solent Transport, an apolitical organisation that brings together all four local councils, is the manager of the MaaS system in operation. Not only does this involve additional, potentially conflicting commercial interests (e.g., with multiple, competing bus companies all needing to make new legal and working relationships with Solent Transport), it also requires the collaboration of four local authorities that may or may not be in the same economic situation or have the same political will to contribute to a scheme that is likely to always need subsidy (Kraus, 2024) (just like public transport more generally; Department for Transport, 2023).

The ActorMap diagrams presented above present visual representations of the stakeholders involved in three different implementations of Mobility as a Service (MaaS) across Europe, all of which are underpinned by software provided by the same MaaS technology provider (i.e., Trafi). The diagrams do not, however, present the links between the actors and organisations. Nevertheless, in developing and discussing the stakeholder landscapes of the three systems with their respective programme managers, and in the subsequent follow-up email communications, details did emerge regarding those relationships.

The most prominent of these was related to funding. In the case of Floya in Brussels, this is an initiative motivated by transport sustainability goals expressed at the national and city levels and carries the implicit acknowledgement that it will not be profitable, rather represents a public service that will always need funding and will continue to be funded (at least in the short to medium term). In Berlin, the Jelbi scheme is also considered part of the public transport offering and funded as a public service; however, there is also the expectation that some aspects of the scheme will generate revenue. For example, the Jelbi stations (represented at the 'Equipment and environment' level in the ActorMap) are physical mobility hub installations and represent a key aspect of the Jelbi business model. The Breeze app, in the Solent region, is currently wholly funded via a UK government research and development initiative; however, it has a goal of being financially self-sustaining without reliance on government subsidy. These funding mechanisms (and goals) are of course related to governance. Therefore, to use Göran Smith's typology (Smith, 2020), Brussels is a public controlled scenario, Berlin a public-private scenario and Solent a market-driven scenario, at least in some respects. This is a major risk to success in the Solent area. Public transport (argued to be the crucial to MaaS, as discussed above) in the UK is heavily subsidised by central government. It is done so as it is considered a public service whose benefits (to society) justify the costs. MaaS appears to be considered in this way in Brussels and Berlin (though perhaps to a slightly lesser extent in the latter). This does not appear the case in the UK. To overcome this risk, greater funding commitments from central government will be required. This presents a challenge given the challenging national economic backdrop.

Implications for ergonomics

The discussion of these aspects arose thanks to the structure provided by the ActorMap and its underlying Risk Management Framework. There exist many different approaches to stakeholder mapping (see Reed & Curzon, 2015 for a review), and we do not claim the ActorMap framework to offer a better or worse approach to doing so (comparative analysis would be required for that);

however, we do argue that the Risk Management Framework, as embodied in the ActorMap template, facilitates a hierarchical understanding of the relevant entities involved or interested in a system (in this case, MaaS), hence is highly useful. The ActorMap also facilitates discussions around communication between actors and the control and feedback mechanisms that exist in the system. One could reasonably argue that the STAMP method (Leveson, 2004) does this to a greater extent; however, resources are not always available to undertake a full STAMP analysis. The ActorMap is a far simpler tool, requiring less time and expertise to complete, yet still offering a hierarchical framework around which to base discussion with system stakeholders.

Rasmussen and Svedung (2000), in their book *Proactive Risk Management in a Dynamic Society*, discuss generic versions of the AcciMap and the ActorMap as tools for, as the book title suggests, *proactively* managing risk in complex systems. We have made the point above that risk (in a system failure sense) does not only relate to safety but also sustainability. A failure to ensure our transport system develops sustainably is a risk to society. We therefore need to be proactive when considering the implementation of interventions aimed at improving transport sustainability.

Although we discuss above the relationships between organisations represented in the ActorMaps, and how information about these relationships emerged in the ActorMap-facilitated discussions with the three MaaS managers, these are not represented on the ActorMap diagrams. Rasmussen and Svedung (2000) discuss role allocation in their chapter on generic ActorMaps; however, inclusion of interconnections is reserved for the AcciMap. Although this can be developed in a generic fashion, its nature lends itself more to events than on-going operations, given the multiple different activities undertaken by each organisation involved (i.e., it would be unwieldy to depict all activities undertaken by all parties in one diagram). As such, we would suggest a development of the ActorMap tool as a tool for stakeholder identification and mapping would be the inclusion of those interconnections. Indeed, our inclusion of colour coding to indicate the nature of each organisation involved is a methodological development of sorts; however, including interconnections would go further. This merits further methodological exploration, although one might reasonably suggest such an inclusion would move this closer to being a STAMP analysis. One of the main benefits of the ActorMap (over STAMP) is its relative simplicity. The extent to which further additions might tip the balance of utility and resource intensity one way or the other will require careful consideration, and likely be dependent on the specific application at hand.

A criticism that could be levelled at the analysis presented above, and at the RMF approach to risk analysis more generally, is that it does not explicitly evaluate risks in terms of the typical impact and likelihood metrics. It also does not have a formalised mechanism of indicating the influence or interest of each identified actor, the dimensions of Mendelow's oft-used stakeholder mapping and engagement matrix (Mendelow, 1981). The ActorMap helps frame discussions, and the template upon which actors are mapped can support a more in-depth understanding of systems hierarchies. Nevertheless, alone it is (arguably) insufficient as a risk analysis, management, and mitigation tool.

Conclusions

The transport sector (and society more broadly) faces a significant sustainability challenge. Systems ergonomics might help us to address that challenge; however, in the human factors and ergonomics literature, system failure has typically been viewed only in terms of safety, not sustainability. We demonstrate the utility of the ActorMap tool, and the Risk Management Framework on which that tool is based, for the identification of risks to the success of a sustainable transport scheme, namely Mobility as a Service. Further research would do well to compare different systems ergonomics methods in this context.

Acknowledgements

This work was funded by the UK's Department for Transport as part of the Solent Future Transport Zone programme.

References

Alyavina, E., Nikitas, A., & Tchouamou Njoya, E. (2020). Mobility as a service and sustainable travel behaviour: A thematic analysis study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *73*, 362-381. https://doi.org/https://doi.org/10.1016/j.trf.2020.07.004

Department for Transport. (2023). *Official Statistics Rail factsheet: 2022*. Retrieved 07/08/2024 from https://www.gov.uk/government/statistics/rail-factsheet-2022/rail-factsheet-2022

Kraus, L. (2024). *Enhancing Mobility as a Service–Sustainability, Willingness to Pay and Profitability Aspects* Dissertation, Duisburg, Essen, Universität Duisburg-Essen, 2024].

Leveson, N. (2004). A new accident model for engineering safer systems. *Safety science*, *42*(4), 237-270.

McIlroy, R. C., Plant, K., Hoque, M. S., Wu, J., Kokwaro, G., Nam, V., & Stanton, N. (2019). Who is responsible for global road safety? A cross-cultural comparison of Actor Maps. *Accident Analysis & Prevention*, *122*, 8-18.

Mendelow, A. L. (1981). Environmental scanning--the impact of the stakeholder concept.

Moray, N. (1995). Ergonomics and the global problems of the twenty-first century. *Ergonomics*, *38*(8), 1691-1707.

Mulley, C., Nelson, J. D., Ho, C., & Hensher, D. A. (2023). MaaS in a regional and rural setting: Recent experience. *Transport Policy*, *133*, 75-85. https://doi.org/https://doi.org/10.1016/j.tranpol.2023.01.014

Rasmussen, J. (1997). Risk management in a dynamic society: a modelling problem. *Safety science*, *27*(2-3), 183-213.

Rasmussen, J., & Svedung, I. (2000). *Proactive risk management in a dynamic society*. Swedish Rescue Services Agency.

Reed, M. S., & Curzon, R. (2015). Stakeholder mapping for the governance of biosecurity: a literature review. *Journal of Integrative Environmental Sciences*, *12*(1), 15-38.

Salmon, P. M., Stanton, N. A., Read, G. J., Walker, G. H., Stevens, N. J., & Hancock, P. A. (2019). From systems ergonomics to global ergonomics: The world as a socio-ecological-technical system. Chartered Institute of Ergonomics and Human Factors Annual Meeting,

Smith, G. (2020). *Making mobility-as-a-service: Towards governance principles and pathways* Chalmers Tekniska Hogskola (Sweden)].

The Age Factor in Ride Comfort: Comparing Younger and Older Passengers' Perspectives

Mikael Johansson, Melina Makris & Anna-Lisa Osvalder

Chalmers University of Technology, Sweden

SUMMARY

This questionnaire study examined the influence of age on factors contributing to overall passenger ride comfort in cars. In total, 1,115 individuals participated, including 269 respondents aged 20–30 years and 260 aged 60–83 years. The questionnaire included two multiple-choice questions about the most and least important factors for front-seat passenger comfort, with eleven factors to choose from. Respondents could provide free-text explanations for their choices. Descriptive statistics quantified selected factors, while thematic analysis explored differences in free-text responses. The results showed that younger and older passengers shared similar preferences, identifying a comfortable seat and feeling safe with the driver as the most important factors. However, preferences differed in the perceived importance of safety features, climate, and functional design, due to variations in life experiences and activity preferences. The findings highlight a shared baseline of factors affecting ride comfort, while also revealing age-specific differences in passenger perception.

KEYWORDS

Ride comfort, car passenger, older adults, younger adults

Introduction

Passenger comfort in cars is crucial, particularly for safety reasons. If the seat and seat belt system cause physical discomfort, such as tensed muscles and static loadings, it may lead to improper use of the protective systems. This increases the risk of injuries during a crash, especially for older adults, who are more vulnerable. Studies have shown that older adults experience greater discomfort from seat belts than younger individuals (Bohman et al., 2019), which may result in misuse of the safety systems (Osvalder et al., 2019). Additionally, the age-related changes in body shape and BMI have shown to impact the seat belt fit and the usage of accessories to mitigate discomfort (Osvalder et al., 2019; Makris et al., 2023a; Makris et al., 2024). Few studies have examined the relationship between car passengers' age and other aspects of physical comfort, such as perceived discomfort due to noise, vibrations or air quality. However, Kwak et al. (2023) studied age-related thermal comfort under different cooling and heating conditions. The results showed that older passengers had different thermal responses compared to younger ones, often feeling less comfortable at equivalent temperature settings. However, comfort is a multidimensional concept that encompasses not only physical, but also psychological and functional factors (Vink & Hallbeck, 2012), which all contribute to the overall ride comfort experience. Functional comfort relates to the ease of use and performance, such as the ability to use the seat and seat belt in a satisfactory manner. Psychological comfort involves aspects such as traveling in a car that feels safe or having a trustworthy driver.

Even though, age-related aspects appear to influence the physical comfort of car passengers, particularly concerning seating ergonomics and seat belt fit, their impact on the overall experience of car ride comfort remains less clear. A questionnaire study was therefore conducted to increase the understanding of passengers' overall ride comfort. The purpose of the analysis presented in this paper was to explore the relationship between age and car ride comfort by comparing the most and least important factors contributing to ride comfort for older and younger front-seat car passengers.

Method

The data was collected through a web-based questionnaire which was distributed to individuals residing in Sweden via an advertisement on the university's social media channels over a three-week period in September 2024. The sole inclusion criterion was that the respondents must have experience travelling in the front passenger seat in a car. The questionnaire comprised two multiple choice questions: "What is most important for you to feel comfortable as a front-seat passenger?" and "What is least important for you to feel comfortable as a front-seat passenger?". For each question, respondents were given the same eleven factors to choose from, covering physical, psychological, and functional aspects of comfort; (i) to be seated comfortably in the seat, (ii) that the seat belt fits comfortably, (iii) to have good visibility out the window, (iv) that the interior is functional so that I can perform the activity I desire (e.g. reach settings and storage), (v) low noise level in the car, (vi) pleasant climate, (vii) feeling safe with the driver, (viii) that the car model feels safe, (ix) few bothersome vibrations, (x) that the interior is spacious so that I have enough space, and (xi) other. Two or three factors could be selected for the first question and one or two factors for the second question. The respondents could also provide free-text answers explaining their choices for the most and least important factors. Demographic data such as age, gender and anthropometrics were also collected.

In the questionnaire analysis, the younger age group was defined as respondents between 20-30 years, while the older age group comprised respondents aged 60-83 years. This classification aligns with Soebarto et al. (2019), who defined older adults as those aged 65 years and above and younger adults as approximately 30 years old to investigate potential age-related psychological differences affecting thermal. Asua et al. (2022) included 15 younger adults with an average age of 29 years and 15 older adults with an average age of 69 years to capture experiences of driving styles and vehicle dynamics to evaluate comfort and motion sickness.

Descriptive statistics were used to quantify the factors that younger and older age groups chose as the most and least important for feeling comfortable as a front-seat passenger. A thematic analysis was conducted on the free-text responses to explore and understand potential differences between the age groups. First, the free-text responses regarding factors that varied in importance between older and younger adults were reviewed and categorized into themes. These themes were then compared to merge similar themes together. Subsequently, the themes of the older group were compared to the themes of the younger group to identify similarities and differences between groups.

Findings

A total of 1,115 individuals answered the questionnaire, with 269 respondents aged 20–30 years and 260 aged 60–83 years. Both groups chose 'a comfortable seat' and 'feeling safe with the driver' as the most important factors (Figure 1). However, the third most important factor differed between the groups, where the older group chose 'the car model feels safe', while the younger group chose 'climate'. Regarding the least important factors, both groups chose 'a functional interior,'few vibrations', and 'low noise levels' as the least important factors (Figure 2). Furthermore, the analysis revealed differences between the age groups' ratings of five comfort factors; 'car model feels safe' - in ratings of both most important (Younger = 16% vs Older = 35%) and least important (Younger = 16% vs Older = 5%) factors, 'climate' - most important (Younger = 42% vs Older = 27%), 'spacious' - most important (Younger = 28% vs Older = 20%), 'low noise level' - least important (Younger = 22% vs Older = 18%), and 'functional interior' - least important (Younger = 40% vs Older = 56%).

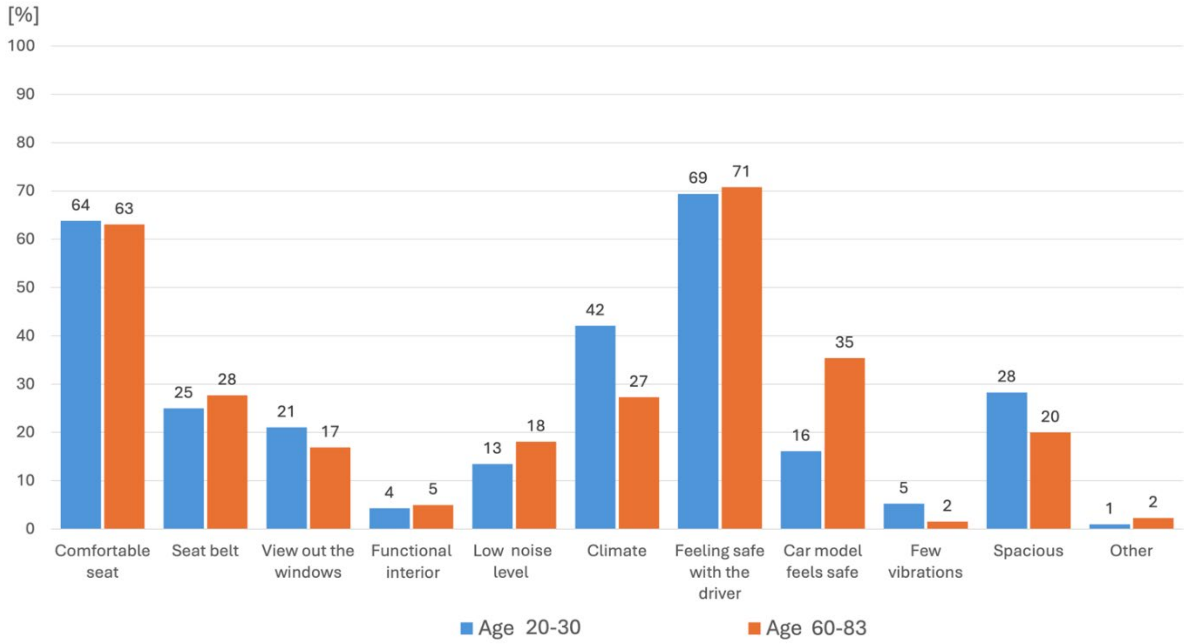


Figure 1. Questionnaire results of the most important factors for feeling comfortable as a front-seat passenger, showing the percentage of respondents in each group who chose the factor as one of the most important.

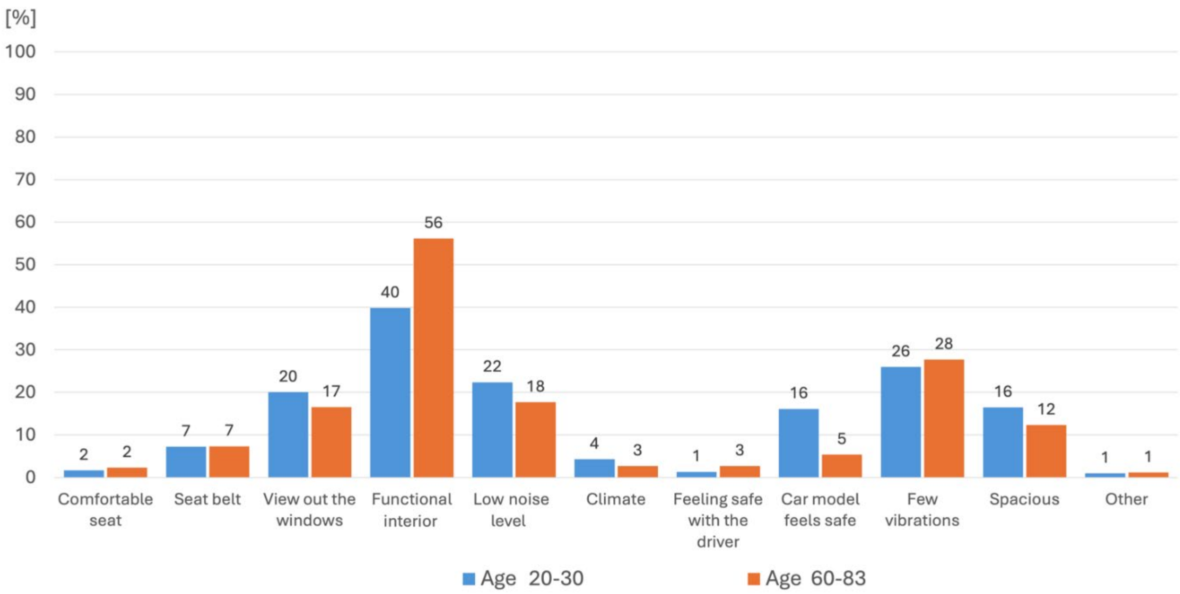


Figure 2. Questionnaire results of the least important factors for feeling comfortable as a front-seat passenger, showing the percentage of respondents in each group who chose the factor as one of the least important.

Similarities and differences in descriptions of comfort factors

For the most important comfort factors, 56% of younger respondents (n=150) and 48% of older respondents (n=126) provided a free-text response explaining their choices. For the least important comfort factors, 47% of younger respondents (n=126) and 36% of older respondents (n=94) also submitted a free-text answer. The analysis of the free text answers showed that in many respects, both older and younger passengers highlighted similar explanations for why they considered certain factors to be the most or least important. However, their descriptions diverged on some points. Regarding 'car model feels safe', which was considered as less important by younger respondents, younger respondents more frequently emphasized specific safety features as essential for feeling secure and relaxed. One respondent explained: "*That the car model feels safe is very important, especially with features like automatic braking if one has been inattentive, as well as ABS and stabilizers."* At the same time, some mentioned safety as an expected baseline standard, for example one respondent stating: *"One thinks that all car models sold should be sufficiently safe so that one does not have to think about it when sitting in the car.*" In contrast, some older respondents referred to their life experiences, such as previous accidents, and noted the importance of a car's safety in reducing the risk of severe injuries during a potential collision. One respondent that referred to a specific previous incident stated that *"I have been in an accident. I survived thanks to the seatbelt",* and another respondent highlighted specific situations where they found it especially important with a safe car: *"It feels calmer with a safe car, especially on roads with a risk of collision with animals or high traffic intensity."*

When describing the reason for choosing 'functional interior', younger respondents more often mentioned technical functionalities, such as infotainment systems, that enable activities during the journey. However, it was also mentioned that such features are often used infrequently, one respondent, for example, described that less significant: *"rarely need to reach for things or handle controls, and besides, you only do it for a short time, so it becomes less important."* Older respondents, on the other hand, more often described that they as passengers rarely perform tasks requiring access to controls, emphasizing instead that functional design is more critical for the driver, one respondent stating that *"the only activity I usually perform during car rides are talking to my fellow passengers and looking out"* and another respondent emphasized that *"If I am a passenger, I should not touch any controls. It is the driver's responsibility."*

When referring to the importance of 'climate', younger respondents often associated it with immediate comfort, focusing on sensations such as avoiding sweating or feeling cold. They placed less emphasis in their responses on potential long-term effects of an unsuitable temperature. For example, one respondent explained that *"It is uncomfortable to be in a car that is too hot or too cold, so having a functioning air conditioning and a comfortable temperature feels important"* and another described that *"I feel nauseous if it is too warm and easily feel cold if it is too cold."* Older respondents, however, tended to incorporate well-being and the long-term impact of climate on the overall travel experience into their reasoning, emphasizing how it affects the journey as a whole. One respondent described that a good climate makes *"The journey becomes comfortable and prevents the body from becoming too stiff... reducing the risk of headaches and motion sickness"* while another stated that climate is important because *"I don't like it when it's drafty, and I don't want to get a backache."*

Discussion

This paper compared the factors contributing to overall ride comfort for older and younger front-seat car passengers. Two factors that stand out are 'a comfortable seat' and 'feeling safe with the driver', which were considered as most important by most respondents in both groups. The importance of feeling safe with the driver has been highlighted in other studies, which state that car passengers tend to feel more comfortable with a safe driver due to several factors related to human psychology and physiology. Safe driving minimizes sudden accelerations, decelerations, and sharp turns, which can reduce the likelihood of motion sickness and discomfort caused by vehicle vibrations (Dacova, 2021). Asua et al. (2022) found that minimizing abrupt maneuvers significantly reduces motion sickness and improves passenger comfort. Additionally, Wang et al. (2020) demonstrated that maintaining a steady speed and smooth driving style significantly enhances passenger comfort by reducing anxiety and providing a sense of security. Furthermore, Mims et al. (2023) highlighted that passengers feel most uncomfortable in situations with distracted drivers or vehicles following too closely, emphasizing the importance of safe driving practices for overall comfort.

Furthermore, both groups pointed out that a comfortable seat is crucial for comfort perception for passengers in cars. A comfortable seat significantly enhances the overall travel experience by reducing fatigue and discomfort, especially during long journeys (Phillips, 2024). Ergonomically designed seats provide proper lumbar support, adjustable features, and high-quality materials that help maintain good posture and minimize strain on the spine. Research has shown that well-designed seats can prevent the onset of musculoskeletal issues and improve passenger well-being by ensuring a stable and supportive seating position (Gracia, 2025). Additionally, comfortable seating contributes to better focus and reduced stress, which are essential for both driver and passenger safety (Robertsen et al., 2022). Makris et al., (2023a) and Makris (2023b) have also discussed the need for ergonomically designed seats and properly fitted seat belts to prevent discomfort and ensure passenger well-being.

Thus, the factors considered most important did not differ substantially. However, this does not necessarily imply that what is considered, for example, a safe driver or a comfortable seat is the same for both groups, even if the factors are rated as equally important. For example, in a study by Osvalder and Bohman (2019) the conclusion was that older passengers require more ergonomic seat and belt designs than younger passengers to accommodate their unique physical characteristics and enhance comfort and safety. The findings revealed that older adults often experienced non-optimal belt fit due to age-related changes in body composition and posture, such as increased kyphosis and altered body shape. This resulted in the shoulder belt being positioned closer to the neck or suprasternal notch, which can cause discomfort and reduce safety.

However, for some factors, the two groups differed in how important or unimportant they considered them. Analysis of the free-text answers showed that while both groups frequently provided similar explanations for why certain factors were considered most important or least important, their reasoning also diverged. The results showed that the older respondents found feeling safe with the car model more important than younger respondents. A possible explanation is that older people have more experience with unsafe cars compared to younger people due to the longer span of their driving history, which includes periods when vehicle safety standards were not as advanced as they are today. Older drivers have witnessed significant changes in automotive safety technology, from the introduction of seat belts and airbags to modern advanced driver-assistance systems (ADAS). This experience can make them more aware of the importance of safety aspects related to the car model. They are also more vulnerable to injuries in the event of a crash due to age-related physical fragility (Ayuso et al. 2019; Alrumaidhi & Rakha, 2022) which may also explain a higher focus on safety.

Another difference was that younger respondents placed more importance on a functional interior compared to older respondents. This can partly be explained by the difference in expectations of what the passenger should be able to do and what functions are necessary to use. Younger participants more frequently highlighted different functions in their explanations, while many older respondents stated that they as passengers rarely perform tasks requiring access to controls. These differences in reasoning may help explain why they prioritized these factors differently, leading to the variation in factors chosen by each group. This emphasises the importance of experiences and expectations in shaping passengers' subjective comfort experience, as also described by Naddeo and colleagues (2015).

Conclusions

The overall conclusion of this questionnaire study is that both younger and older respondents share similar preferences regarding the factors that contribute most (feeling safe with the driver and sitting comfortably in the seat) and least (functional interior and few vibrations) to overall passenger ride comfort in cars. This suggests that there is a baseline set of factors considered particularly important for front-seat passengers, regardless of age. However, notable differences were observed in the perceived importance of certain factors, particularly the significance of feeling safe with the car model, where older respondents found this a more important factor than younger ones. These differences can partly be explained by variations in experience over time as car passengers and the activities respondents wish to engage in as passengers.

References

Alrumaidhi, M., & Rakha, H. A. (2022). Factors Affecting Crash Severity among Elderly Drivers: A Multilevel Ordinal Logistic Regression Approach. *Sustainability*, 14(18), 11543. Retrieved from https://doi.org/10.3390/su141811543

Asua, E., Gutiérrez-Zaballa, J., Mata-Carballeira, Ó., Ruiz, J. A., & del Campo, I. (2022). Analysis of the Motion Sickness and the Lack of Comfort in Car Passengers. Applied Sciences, 12(8), 3717. https://doi.org/10.3390/app12083717

Ayuso, M., Sánchez-Reyes, R., & Santolino, M. (2019). Does longevity impact the severity of traffic accidents? A comparative study of young-older and old-older drivers. *Institute of Applied Economics (IREA) Working Papers*. Retrieved from https://www.ub.edu/irea/working_papers/2019/201908.pdf

Bohman, K., Osvalder, A-L, Ankartoft, R. & Alfredsson, S. (2019). A comparison of seat belt fit and comfort experience between elderly and younger front seat passengers in cars. Traffic Injury Prevention, Vol 20, 1538-9588 (ISSN) 1538-957X (eISSN).

Dacova, D. (2021). Ride comfort in road vehicles: a literature review. *International Scientific Journal Trans & MOTAUTO World*, 2, 65-72. Retrieved from https://stumejournals.com/journals/tm/2021/2/65

Gracia, A. (2025). What Is a Coupe Car? Exploring Different Coupe Styles. Retrieved from https://detailedvehiclehistory.com/blog/what-is-a-coupe-car

Kwak, J., Chun, C., Park, J.-S., Kim, S., & Seo, S. (2023). The gender and age differences in the passengers' thermal comfort during cooling and heating conditions in vehicles. *PLOS ONE*, 18(1), e0294027. https://doi.org/10.1371/journal.pone.0294027

Makris, M., Bohman, K., & Osvalder, A. L. (2023a). Comparison of sitting postures and shoulder belt fit of rear seat car passengers over time in stationary and driven scenarios. In *Proceedings of the IRCOBI Conference, Cambridge* (pp. 690-707).

Makris, M. (2023b). *How Does it Feel and How is it Measured? Assessing Sitting Comfort and Postures of Rear-Seated Car Passengers in Stationary and Driven Scenarios Over Time*. Chalmers Tekniska Hogskola (Sweden).

Makris, M., Osvalder, A.-L., Bohman, K. (2024). Comfort experience of rear seat car passengers over time in stationary and driven scenarios. Manuscript submitted to Transportation Research Interdisciplinary Perspectives.

Naddeo, A., Califano, R., Cappetti, N., & Vallone, M. (2015). The effect of external and environmental factors on perceived comfort: the car-seat experience. Proceedings of the Human Factors and Ergonomics Society Europe, 291-308.

Osvalder, A-L., Bohman, K. et al (2019). Seat Belt Fit and Comfort for Older Adult Front Seat Passengers in Cars. Scientific article, 2019 IRCOBI Conference Proceedings - International Research Council on the Biomechanics of Injury, IRC-19-12, Florens, Italy, 11-13 Sept.

Phillips, C. (2024). Understanding Car Ergonomics: How Design Impacts Comfort and Safety. *ErgoHealth Solutions*. Retrieved from https://www.ergohealthsolutions.com/post/understanding-car-ergonomics-how-design-impacts-comfort-and-safety

Robertsen, R., Lorås, H. W., Polman, R., Simsekoglu, O., & Sigmundsson, H. (2022). Aging and Driving: A Comparison of Driving Performance Between Older and Younger Drivers in an On-Road Driving Test. *SAGE Open*, 12(2), 1-9. Retrieved from https://doi.org/10.1177/21582440221096133

Soebarto, V., Zhang, H., & Schiavon, S. (2019). A thermal comfort environmental chamber study of older and younger people. *Building and Environment*, 156, 1-10. https://doi.org/10.1016/j.buildenv.2019.03.032

Vink, P., & Hallbeck, S. (2012). Comfort and discomfort studies demonstrate the need for a new model. In (Vol. 43, pp. 271-276): Elsevier.

Using Co-Simulation to Explore Distributed Situation Awareness in AV Remote Operation

Hannah Parr, Catherine Harvey, David R. Large & Sarah Sharples

Human Factors Research Group, University of Nottingham, UK

SUMMARY

We introduce a highly novel, multi-participant co-simulation study to explore distributed situation awareness between a remote operator, a connected and automated vehicle (CAV) system and CAV user in two highly automated vehicle use cases. Initial reflections on the success of this method are reported, as well as plans for further analysis.

KEYWORDS

Distributed Situation Awareness, Automated Vehicles, Remote Operation

Introduction

Even after the widescale introduction and adoption of high-level connected and automated vehicles (CAVs) at SAE Levels 4 and 5 (SAE International, 2021), issues may still occur where the automated driving system (ADS) encounters unpredicted events which cannot be resolved by the vehicle technology alone. These 'edge-cases' will cause issues for the operation of CAVs and will likely require human intervention at some level to resolve. Remote operation has been proposed to provide this human support and this will encompass many different tasks or 'competencies'. Parr et al. (2024) identified four types of remote operation in this context: remote monitoring (RMo), remote assistance (RA), remote management (RMa) and remote driving (RD), which differ in the expected types of tasks, competencies and interactions between the CAV, CAV user and remote operator (RO), and may depend on the specific use-case under consideration. To be able to intervene in an appropriate and timely manner, the RO must quickly build appropriate situation awareness (SA). However, as CAVs are out of the direct line of sight of an RO, operators will require novel interaction methods to interact with the vehicle and CAV user/s. Multiple 'agents' (e.g. other road users, emergency services personnel, technological systems, etc.) may also be involved in an interaction, highlighting the need for successful teamworking, underpinned by distributed SA (i.e. where each individual agent possesses unique elements of awareness which are compatible and collectively required (Stanton et al., 2013)).

To define the information requirements of future ROs and CAV users, it is important to understand how interactions between ROs and other agents (particularly CAV users) occur in realistic CAV use cases. The aim of the current study was therefore to develop a remote operation co-simulator to enable naturalistic interactions between participants who were invited to take the roles of RO and CAV user (with behaviours of other system agents being controlled via the simulation). This two-participant design allowed interactions between the RO and vehicle occupant to emerge naturally, which we believe would have been inhibited if one of the roles was experimenter-controlled. Three simulation scenarios were designed to encourage the sharing of information between participants in order to collaboratively manage the unfolding situation in each case. To the best of our knowledge this is the first study to use naturalistic, multi-participant simulation to examine the roles of both a RO and a CAV user within a CAV context.

Co-Simulator Design

The CAV user participant was seated in the front passenger seat of a driving simulator, which was designed to simulate a driverless taxi (with primary controls hidden from view), shown in **Figure 1**. A screen on the centre console displayed destination and journey progress and provided audio alerts and information during the remote operation events. Two cameras were placed in the vehicle, to film the forward view and the passenger, enabling a real time video view of the incident and passenger for the RO. AV simulation software (SCANeR) was used to create three scenarios, each depicting a different use case (informed by Parr et al, 2024), and covered a Practice Event: Vehicle software failure, Event 1: Path disruption, Event 2: Vehicle-vehicle collision. In each scenario, the vehicle travelled autonomously through different road environments (rural/residential/city) for 5-7 minutes before encountering a ‘critical’ event designed to require intervention from a RO, necessitating RMo, RA, and RMa (but not RD) actions.

Figure 1: Photo showing the simulator vehicle set-up used during the study. The driving mechanisms (steering wheels, pedals, gearstick and handbrake) are hidden from the participant seated in the front passenger seat.

The RO was located in a separate laboratory space and provided with a workstation comprising five screens. These included a central control display, a display showing live video of the AV interior and forward view (during remote operation only), a task monitoring display and two additional information displays, shown in Figure 2. Before receiving a request to intervene, the RO participant engaged with a basic monitoring task, intending to mimic expected activities associated with RMo. An audio communication link between AV user and RO was activated during critical events. Standard Operating Procedures (SOPs) were developed from previous work (Parr et al., 2024) and provided as guidance for the RO using ‘yes-no’ flowcharts with suggested actions.

Figure 2: Photo showing the Remote Operation Workstation during a remote operation event (no participant seated in the simulator vehicle), showing on the screens (Top, left to right) weather updates and a generic rolling information screen and (Bottom, left to right) the remote monitoring task, the main remote operation interface and the link to the live view of the road and the user.

The Study

Twenty-two participants took part (13M, 9F; age range 21-70), with two participants per study. One participant was assigned the role of AV User and the other the role of RO; participants did not know each other previously. All ROs had a valid UK driving licence. Both participants received the same ‘study overview’ information in advance of the study and this was re-iterated at the beginning of the study. The RO participant was familiarised with the capabilities of the remote operation workstation, before both participants engaged in a practice event, which simulated a vehicle software failure. Following this, participants could ask questions and clarify issues with the experimenter. Participants then took part in two study scenarios (1 and 2, above) which were audio and video recorded. Prior to each ‘critical’ event, the AV user was encouraged to behave as they would during a taxi journey and the RO participant was asked to imagine they work for an AV taxi company and were required to monitor and respond accordingly. When a ‘critical’ event occurred, the AV user was provided with details on the in-car screen plus audio prompts. The RO received notification of an intervention request, and the remote operation interface became active, enabling 2-way audio communication between the RO and the vehicle cabin. Participants were instructed to attempt to resolve the critical event, with the goal of allowing the AV user to safely continue their journey. Post-event, both participants completed a SART questionnaire (Taylor, 2017) and were asked some general questions about their experience. This was followed by a Critical Decision Method (CDM; Klein et al., 1989) interview to explore behaviours and attitudes with both participants present, along with some additional unstructured interview questions to explore any concepts and attitudes addressed by the participants that were not easily explored by the CDM probes.

Reflections on Method and Initial Findings

The data analysis was completed in two stages; firstly, the perceptual cycle model (PCM) was used to map information transfer between agents and the elements identified through the CDM interview to evaluate the presence of distributed situation awareness (DSA), and the benefit of using this theory of situation awareness (SA) when considering the skills and tasks of the RO. The SART data and eye-tracking was used to give further context to the collected data. Secondly, a thematic analysis was conducted on the event transcripts from the simulation and interview transcripts to investigate emerging themes. In conjunction with SART rating, results will be used to develop and validate the operator sequence diagrams in Parr et al. (2024). Early observations suggested participants felt engaged in the simulated environment and believed they were influencing the simulator actions. RO participants were able to take on the technical role of remote operation following the provided training and by using the SOPs. Specific comments and themes raised in the thematic analysis suggest that CAV users found the live audio link to the RO increased trust and reduced feelings of isolation, compared to only having pre-programmed audio prompts. Issues of abuse and misuse were also raised alongside the concept of the CAV User individual agency during remote operation. RO participants highlighted the importance of corroborating information from several sources including the CAV user, especially when dealing with issues of software failure. Further analysis explored distributed SA between agents, information requirements of RO and vehicle occupants, and competencies required for RMo, RA, and RMa roles, to support the design of RO workstations and working practices. The importance of ensuring the CAV User is 'in-the-loop' with the RO was highlighted, as their actions, world information and planning may heavily influence the ROs further action and examples of inaccurate schema, information sharing and the need to communicate with other scenario agents were highlighted.

Conclusion

Through this study, we introduced a highly novel, multi-participant co-simulation study to explore the interactions between a remote operator, a connected and automated vehicle (CAV) system and CAV user in two highly automated vehicle use cases which act as exemplar scenarios for the application of remote operation. The analysis of this work focused on establishing the importance of DSA, through the application of the CDM and subsequent PCM analysis, followed by an inductive thematic analysis to explore emerging themes present in the study and enable recommendations for the implementation of remote operation into a commercial service.

References

Klein, G. A., Calderwood, R., & MacGregor, D. (1989). Critical Decision Method for Eliciting Knowledge. IEEE Transactions on Systems, Man, and Cybernetics, 19(3), 462-472. https://doi.org/10.1109/21.31053

Parr, H., Harvey, C., Burnett, G., & Sharples, S. (2024). Investigating levels of remote operation in high-level on-road autonomous vehicles using operator sequence diagrams. Cognition, Technology & Work 26(2), 207-223. https://doi.org/10.1007/s10111-024-00762-w

SAE International (2021). 3016_202104 Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. https://www.sae.org/standards/content/j3016_202104/

Stanton, N. A., Salmon, P. M., Rafferty, L. A., Walker, G. H., Baber, C., & Jenkins, D. P. (2013). Human Factors Methods: A Practical Guide for Engineering and Design (2nd ed.). Taylor & Francis Group

Taylor, R.M., 2017. Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In Situational awareness (pp. 111-128). Routledge.

Using Design with Intent to Encourage Active Travel in Mobility as a Service

Joy McKay, John Preston

University of Southampton

SUMMARY

Mobility as a Service (MaaS) apps have been developed in the hope of increasing sustainable travel, yet the most sustainable forms of travel; active travel, are yet to receive substantial consideration. This significant change of increasing walking and cycling requires users to have all of their information needs met. Using COM-B alongside Design With Intent users, will co-create in-app features intended to provide the assistance needed to effectively increase active travel use.

KEYWORDS

MaaS, Active Travel, DWI

Introduction

Mobility as a Service (MaaS) apps are mobile phone applications offering a 'one stop shop' which aim to enable the user to Plan, Book, Pay and Navigate journeys which could be completed using a variety of travel modes, such as public transport (such as buses and trains), micromobility (such as hired escooters, bikes or ebikes), active travel (such as walking, running or own bike use), on demand services (such as taxis), own car use and use of more traditional hire vehicles such as can or van hire schemes. By presenting all of the available options to the user the hope is that they will have the information to make more sustainable and healthy choices about how to make a journey. The ultimate aim is that MaaS will encourage behaviour change, reducing the reliance on the private car, particularly for solo, short to medium length trips (Richardson et al, 2022).

The Solent Future Transport Zone (FTZ) was one of four FTZ schemes awarded funding from the Department for Transport to fund such an MaaS app and work commenced in 2020. The project has been managed by Solent Transport, a group combining Southampton City Council, Portsmouth City Council, Hampshire County Council and Isle of Wight Council. The Human Factors Engineering Team, part of the Transportation Research Group at The University of Southampton has been commissioned to optimise the interface as part of a user-centred design process.

The importance of including active travel modes as options in MaaS is of importance in the UK government's plans for MaaS. Prior to the awarding of the FTZs the 'Future of Mobility: Urban Strategy' plan was released. It includes nine principles including that "Walking and cycling must remain the best options for short urban journeys" (p8. gov.uk 2019) further stating that "Mobility as a Service could… move people towards active and sustainable modes" (p27. gov.uk. 2019). Subsequently, the Mobility as a Service: Code of Practice was published. In it there are several suggestions for what information may help choose to use active travel such as high-quality maps, information on CO_2 savings and the associated health benefits of each trip (gov.uk 2023).

The logistical considerations involved in making all public transport providers and ticket types available in one app (particularly complicated in an area under the control of multiple local authorities and with several bus and train companies), combined with the financial benefit of public transport, micromobility and car hire providing revenue from in-app sales, have meant that the provision of support and information relating to active travel is still yet to be realised.

Method

The COM-B Model of behaviour change has been used extensively in the fields of public policy and health. It suggests a simplified model of three factors to consider when attempting to influence behaviour change: Capability, Opportunity, and Motivation, (Michie et al, 2011). Therefore, a user in possession of all of these factors is likely to be more successful in making the desired changes (Cane et al., 2012). As Human Factors practitioners, we wish to ensure that the MaaS app can offer genuine impact to the public by addressing many of these factors when developing new features, from a user-centred perspective.

Design with Intent (DWI) is a user centred design method whereby workshops with participants from target end user groups are conducted in order to generate novel design ideas with the aim of influencing a particular behaviour (Lockton et al, 2012). During the initial design phase DWI workshops were conducted in order to generate novel ideas from members of the public for features which could be included in this new MaaS app (Kim et al., 2023).

This proved to be a useful exercise for developing prototypes and therefore it was decided that a subsequent set of DWI be conducted but to focus on active travel and with knowledge of what the app offers currently, and what competitor or complementary products are able to offer. Participants will be those who wish to either start or increase their amount of active travel and will choose to join a group for either walkers or cyclists. Participants will then design one or more in-app features which they believe would help them in their wish to use active travel. They will be encouraged to think about what knowledge they may need to possess (capability) what education or environmental needs they may require (opportunity) and what could incentivise these changes (motivation).

Results

The ideas generated from these workshops will be developed into further prototypes for delivery to the project partners, forming part of a report on design recommendations focusing on active travel, but also published with the intention that these could form part of the growing literature on MaaS.

References

Cane, J., O'Connor, D. & Michie, S. Validation of the theoretical domains framework for use in behaviour change and implementation research. *Implementation Sci* **7**, 37 (2012). https://doi.org/10.1186/1748-5908-7-37

Gov.uk (2019) Future of Mobility: Urban Strategy. https://assets.publishing.service.gov.uk/media/5dcd8417ed915d071ca239e9/future-of-mobility-strategy.pdf

Gov.uk (2023) Mobility as a Service: Code of Practice https://www.gov.uk/government/publications/mobility-as-a-service-maas-code-of-practice/mobility-as-a-service-code-of-practice#enabling-active-and-sustainable-travel-1

Kim, J., Howarth, H., Richardson, J., Preston, J. (2023). User-centred generation of early-concept Mobility-as-a-Service interface designs aimed at promoting greener travel. In: Tareq Ahram, Waldemar Karwowski, Pepetto Di Bucchianico, Redha Taiar, Luca Casarotto and Pietro Costa (eds) Intelligent Human Systems Integration (IHSI 2023): Integrating People and Intelligent Systems. AHFE (2023) International Conference. AHFE Open Access, vol 69. AHFE International, USA. Http://doi.org/10.54941/ahfe1002884

Lockton, D., Harrison, D. & Stanton N. A. (2012) Design With Intent 101 Patterns for Influencing Behaviour Through Design http://imaginari.es/wp-content/uploads/2023/04/designwithintent_cards_1.0_draft_rev_sm.pdf

Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci. 2011 Apr 23;6:42. doi: 10.1186/1748-5908-6-42. PMID: 21513547; PMCID: PMC3096582.

Richardson, J., Howarth, H. & Kim, J. (2022) Developing a Heuristic Tool for Evaluation of Mobility as a Service (MaaS) Interfaces. In Ergonomics and Human Factors 2022. Balfe, N. & Golightly, D. CIEHF

Water Quality Investigations and Root Cause Analysis: A Human Factors approach

Tracey Milne & Jodie Dix

IHF Ltd

SUMMARY

This project aimed to integrate Human Factors (HF) into water quality investigations to enhance understanding of human behaviour in incidents, improve coordination, and ensure better system management. By developing an HF Root Cause Analysis tool, the client identified gaps in their investigation process, leading to a revised approach that fosters a learning-driven, human-focused culture for water quality events.

KEYWORDS

Utilities, root cause analysis

Introduction

The client (a large utilities company) wanted to understand how the recognition of Human Factors elements could ensure comprehensive understanding and effective management of water quality events.

By considering Human Factors, they would acknowledge the role of human behaviour, cognition and communication in water quality incidents and then create strategies that would address both technical and human elements to enhance water quality systems, improve response coordination and contribute to the safeguarding of water resources and public health.

The aims of the project were to:

- Integrate HF into incident investigation processes.
- Improve investigation outputs.
- Leads to better results and actions.
- Implement and train.
- Create a safer and more productive working environment.
- A reduction in the number of accidents.

Method

An agile project (consisting of five sprints over 15 weeks) was instigated to explore how Human Factors could be integrated into water quality investigations.

Sprints are fixed length projects of work which ensure short iterations for feedback to inspect and adapt both how the work is done and what is being worked on. This project consisted of five sprints:

The first sprint was a workshop which focused on organisation readiness by understanding the current water quality process.

The second sprint focused on a document review, as well as understanding roles, commitments and expectations of the project and looking at what is and not working currently within the organisation.

The third sprint reviewed interventions and introduced the SUPA model.

The fourth print was a workshop that set out to evaluate the proposed solutions.

The fifth and final sprint was a retrospective review of the project and the production of a draft project report.

Information was collected by reviewing documentation, site visits, engagement workshops and feedback analysis.

Results

The initial stages of the project revealed little evidence of Human Factors being integrated and applied within the water quality investigation process. It was apparent that the overall investigation process was complex and required input from a multidisciplinary team of stakeholders.

It was identified that although the investigation process was thorough, there was a preoccupation of technical system characteristics over the human's making alterations to the water quality systems and the justifications for doing so.

IHF created the HF Root Cause Analysis tool in collaboration with the client the address Human Factors shortcomings within the investigation process. This tool was able to identify human error types, uncover critical moments in an incident and ascribe Performance Influence Factors to an event – recognised as essential Human Factors functions in an investigation.

Key takeaways

The development of this tool allowed the client to identify gaps in knowledge and competency concerning the application of Human Factors knowledge when using the tool. The client subsequently revised their investigation process to include stakeholders across the entire investigation process and allow those involved to be included in generating meaningful recommendations.

The client is now committed to being a Human Factors oriented organisation with a focus on creating a learning driven and just and fair culture which recognises the human contribution to water quality events.

Author index