Chartered Institute of Ergonomics & Human Factors

Contemporary Ergonomics & Human Factors 2023

Editors: Dave Golighty, Nora Balfe & Rebecca Charles

Contemporary Ergonomics & Human Factors 2023

Editors

Dave Golightly, Nora Balfe & Rebecca Charles

Chartered Institute of Ergonomics & Human Factors

ISBN: 978-1-9996527-5-3

Preface

Contemporary Ergonomics and Human Factors 2023 contains the proceedings for the Chartered Institute of Ergonomics & Human Factors Annual Conference, which was held 25-26 April 2023 in Kenilworth, UK.

We invited all authors to submit short, two-page papers. This provided a platform to put forward experimental research, discipline reflections and applied case studies. We had an overwhelming response to the call and received almost 100 submissions. Many of these reported on ongoing work or early-stage research, which is not always well represented at academic conferences. The papers were all peer-reviewed by our Programme Committee and many authors were then invited to submit long papers.

As always, we received extremely high-quality submissions, which were varied in content and application. Themes included automation, physiological measures, design of equipment, assessment of new technology, safety culture, accident investigation, human factors methods and artificial intelligence. Application areas included transport, healthcare, manufacturing, construction, nuclear and defence.

We saw a great number of submissions from the healthcare domain, covering various topics such as equipment, device and information design, using human factors techniques to evaluate and improve existing healthcare systems. Also included were areas which have become more prevalent over the past few years, such as telephone triage and procurement of medical supplies.

We believe that the papers represent the state-of-the-art in ergonomics and human factors, on an international level. 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 2023 Programme Committee

Dave Golightly (Co-Chair)
Nora Balfe (Co-Chair)
Rebecca Charles (Co-Chair)
David Caple
Dipak Chauhan
Enda Fallon
Richard Farry
Ella-Mae Hubbard
Steven Mallam
Celine McKeown
Andrew Thatcher

CIEHF Organising Team

Tina Worthy, Chief Operating Officer
Rachel Stalker, Submissions Administrator

Contents

Modelling User Contribution to Capability Within a Supervisory Control System

Mike Tainsh

BAE Systems Maritime

ABSTRACT

Supervisory control is a common category of system employed for many surveillance applications and is a continuing subject of interest to ergonomists. During their development, following an initial statement of system requirements, capability options need to be assessed to understand the contribution of design features to system effectiveness. One technique that can be employed is capability modelling which aims to generate predictions of outcomes dependent on initiating events. A novel capability modelling technique is proposed based on an integration of ergonomics research results and professional input.

KEYWORDS

System development, capability modelling, supervisory control

Introduction

Over recent years there has been a trend to develop supervisory systems, where the person carrying out the controlling and the object being controlled become ever more separated and the controlling process more complex. Previous maritime studies have addressed some of the general aspects this topic (Tainsh, 1982). However, there is little ergonomics modelling work to address the broader systems design issues that are identified early in the development cycle when we work with our engineering colleagues to develop design options. This modelling work could include capability modelling (Lindbom, Tehler, Eriksson and Aven, 2015).

Hence, it is useful to briefly review the results from research studies to support the development of a modelling technique to support capability studies of future systems.

This work focuses on modelling at the early stages of the development process before the start of design work. It addresses system capability and system architecture and capability. The systems engineering aim is to model capability which leads on to system architecture and design. The foci of this work are users with their equipment, and the means of achieving predicted goals.

It appears common for work on capability to take account of the organisational issues and hence this approach provides a useful starting point for ergonomics.

The aim of this study is to model the user characteristics of future capability of supervisory systems prior to the architectural stage of the development process.

Initial Considerations of Development and Architecture

The starting point is an approach based on the development process as given in ISO 15288 (2013).

In ergonomics, the use of a layered functional model of user characteristics has been referred to as a User System Architecture (USA) (Tainsh, 2018). This functional description provides a framework which can be populated by human/organisational detail of the roles, tasks and the activities to be carried out.

The issues to be addressed initially are:

1. Execution of Capability:
   (a) The system capability should match the demand required to a level that meets the risk requirements.
   (b) The human contribution must match the contributions of the other parts of the system, its hardware, functionality or other features.
2. The characteristics of the control process. The progress of the controlling activities needs to be understood so that it matches the required operational timescale.
3. The team organisation. The team members who may be remote from each other and the items being controlled.

Literature Review

Lindbom et al address the issues associated with preparing a system for uncertain outcomes. They have provided a definition of capability:

> "A description of capability based on our definition includes descriptions of the initiating event, the performed task, the consequences associated with the performed task, the uncertainties concerning these consequences and the background knowledge, which form the basis for these descriptions."

Capability is associated with the consequences of the system's operation. We wish to model the human components of the capability to indicate the potential of the system to manage uncertain events. This work suggests that we employ a concept of capability with a consideration of roles, organisation and resources to link with a concept of risk. It provides a set of concepts from and engineering background which enable us to link ergonomics issues into a broader disciplinary framework. Risk is defined in terms of uncertainty of outcome and the severity of the consequences.

Large scale systems and levels of automation

Sheridan (1983) specified the characteristics of supervisory control which included operators working remotely from the object or events under their control. He characterised the control tasks as having two components initialisation and performance. He provided a detailed set of considerations of the advantages and disadvantages of various design strategies.

Supervisory Control

Sheridan and Hennessy (Workshop 1984) characterised a supervisory control system and the tasks that could be carried out by it. The importance of the user trusting that the system will carry out its allocated functions was emphasised, along with the complementing levels and where control was located.

Management of Multiple Dynamic Human Supervisory Control Tasks

Mitchell and Cummings (Workshop 2005) linked the concepts associated with levels of automation with workload and the damaging consequences of overloading. This is placed in a context of information waiting to be handled so that the time taken to switch from one object to another may result in undesired consequences.

Team Behaviour

Artman (2000) investigated team behaviour and its dependence on the allocation of tasks within the team. In particular, it is important whether team are carried out in serial or in parallel as there may be “overheads” in the amount of communication within the team as a consequence.

Ecological displays

Burns (1999) investigated ecological displays to understand the design criteria and parameters to ensure effectiveness. In the course of this work, it was possible to investigate the characteristics of the users when visual scanning. This is important as it yields evidence of how users perform the task of understanding the situation that they are attempting to control.

Initial considerations

The following topics need to be addressed and variables represented so that they can be included within the modelling process:

- The roles/tasks/events that enable the goal to be achieved. The set of roles, tasks and events that can be modelled depends on the performance data available. It appears that much of this data is held within proprietary databases.
- The acquisition, collation and integration of information are likely to be component tasks of any but the simplest of control systems and are very likely to be included within a system with multiple sensors and users.
- Time-based performance characteristics. In the MoD UK, maritime world, we have the benefit of work carried out over 40 years ago at the EMI Laboratory for the MoD. The validity of this performance data underpins the value of this technique.

- Performance of controller and the exercise of control via a control loop is likely to be an important focus of any investigation of supervisory activities, to understand how performance on decision-making may depend on system design - including the implementation and use of automated functions.

The Approach

The highest level statements within the USA will be expressed in operational and systems engineering terms. For ergonomics practitioners with systems engineering, a User Systems Architecture (USA) provides a framework (i.e., a structure with a set of design constraints) that provides a starting point for the development of the representation.

A generic layered description for a supervisory system working for surveillance purposes is given in Table 1. This applies to both individual users and a team working together.

Table 1: A description of the USA layers

Layer number	Layer Name	Brief description
1	User goal, scenario/ context and constraints	A description of a context in which the system operates
2	Business description	The business processes which satisfy the scenarios
3	A technical description	Overall technical system structures, forms and processes
4	Users' roles and team organisation	An ergonomics description of the team its organisation and individual roles.
5	Individual User's characteristics and tasks	Each participants tasks, activities and personal descriptions

The approach to capability modelling

The organisational arrangements (Figure 1) to support the control of the surveillance system are similar to those previously described in this set of investigations (Tainsh; 2018).

The approach has been:

- The first stage is to verify the description of information associated within the USA and the implied control loop as described in Figure 1. While there may be maritime traditions that have to be taken into account, the use of automation is likely to be widespread even when unnoticed by the users. The work of Sheridan is useful to help understand the characteristics of the USA.
- The control loop(s) are central portions of the modelling process, and it is critical to ensure a full understanding of this loop to enable a valid investigation which includes the assignment of performance values to the activities and tasks.
- The assignment of performance values is associated with at least two major difficulties dependent on the processes which are being controlled:
  (a) learning;
  (b) boredom and fatigue.
- The assignment of values to the performance characteristics needs to be agreed/validated in discussion with user representatives.

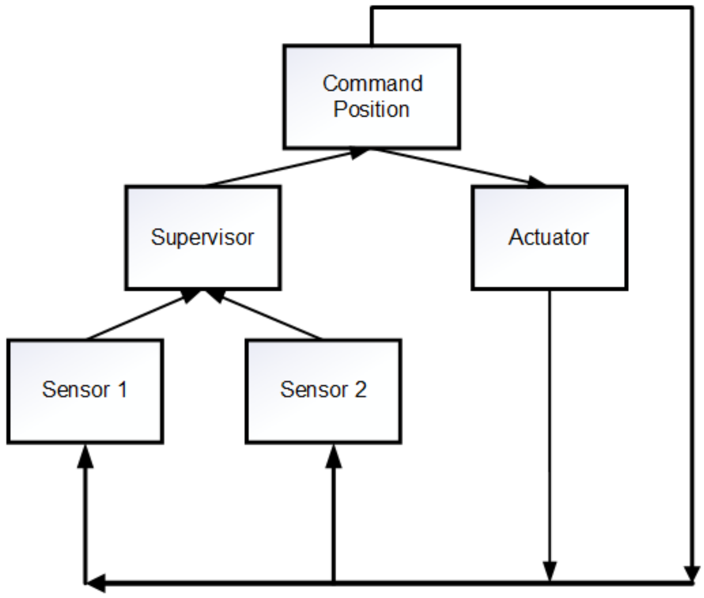


Figure 1: Control diagram for supervisory system showing roles and information flows

Design criteria for this model

Investigation of Resource Demands. For many years, there has been a concern in system design that users were likely to be overwhelmed by the rate of data coming into a control system. However, there are no techniques to help understand the extent of the problem or its mitigation. The model must be open to the investigation of a variety of demands.

Investigation of Automation and Complement. The introduction of automated functions is sometimes unnoticed for example the user of algorithmic techniques for signal processing would often never be considered as automation by ergonomists. However, there are many cases where automation impacts directly on the way that users carry out control tasks and the system effectiveness. Hence the technique must be able to investigate alternative automation techniques.

Investigation of Potential System effectiveness. For many maritime systems, the prime requirement is to be effective, and that means enabling the system and the users within it to carry out control actions as swiftly and accurately as possible. This technique is required to indicate that time-based schedule of the outcome of its controlling actions.

A Generic Model

The current model has been implemented in an Excel spreadsheet using the standard Excel functions and the information provided above. Initiating events are specified. There are limitations when using a spreadsheet in some applications but in this case the accuracy of the calculations was estimated to be in line with the accuracy and precision of the values of the performance variables.

The generic characteristics of the user components of the model are shown in Figure 2:

- Process 1 involves detection and includes initial assessment which enables the object being controlled to be moved to the next process i.e Process 2. The learning here will involve knowing that an object has been detected.
- Process 2 will involve the main controlling processes and involve understanding the classification of the object and the necessary controlling activities.
- Process 3 is included here to include the handling of objects which involve high levels of risk and need to be handled differently from those handled in Process 2.

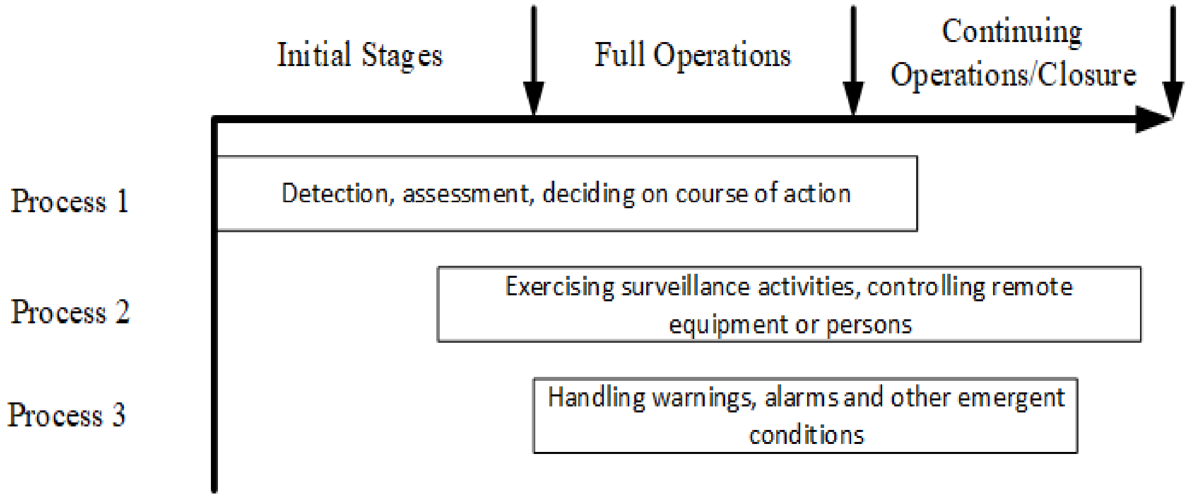


Figure 2: Timeline for Operations

Results

The table of technical information used in this study is given in Table 2.

Task models are derived for each user role, the sets of tasks and activities as specified within the USA. In each case the role, task and activities model have similar characteristics to that shown in Figure 2. Clearly these will depend on the application and requirements but in the maritime applications investigated up until now, there appears to be a common pattern.

Table 2: Technical information used in within the modelling technique

Descriptor	**Range of values**	**Reference**
Role and Task design	Goal dependent	Sheridan, T. B. (1983) and Sheridan, T. B. et al (1984)
Scan times	Within the range 15 – 120 secs	Burns, C. M. (1999), Mitchell P. J. et al (2005)
Performance times for mouse with flat screen	Detection 5 secs, Appreciation 5 secs, Selection/Deselection 5 secs	EMI Electronics (1979)
Performance Shaping Factors for learning	Within the range 0.25 to 1.0	Experience with Users
Allocation of resources at individual and team levels	Within the range 0.1 to 1.0	Experience with Users
Effectiveness and safety criteria	Scenario dependent	In agreement with Users
Team Organisation	Serial versus parallel organisation	Artman, H. (2000)
Supporting equipment functionality	Equipment dependent	In agreement with system engineering specialists
System architectural options	System dependent.	In agreement with system architectural specialists

The results for the whole control team can be aggregated as shown in Figure 3. This shows the performance of one team (successes and failures) in the event of three possible sets of initiating events: low, medium and high demands. In each case, it was seen that user performance starts at a relatively low level as the tasks develop and then both success and failure will change dependent on the characteristics of the role, task and activities. With low levels of demand the control achieved must be balance against potential failures but given time the situation appears to come under control. In the case of medium demand the situation may be brought under control but the time taken may be unacceptable.

In the case of high levels of demand the situation incurs the possibility of moving out of control with an unacceptably high level of failure.

The design aim is directed towards understanding the capability required to handle normal, extreme or other sets of circumstances that may define demand.

The use of figures such as this show performance against requirements. In particular, it can be seen that for this team as the demand associated with the initiating event(s) increases, the risk of failure increases. Hence, we can estimate the likely maximum performance for varying levels of demand and help understand the risk.

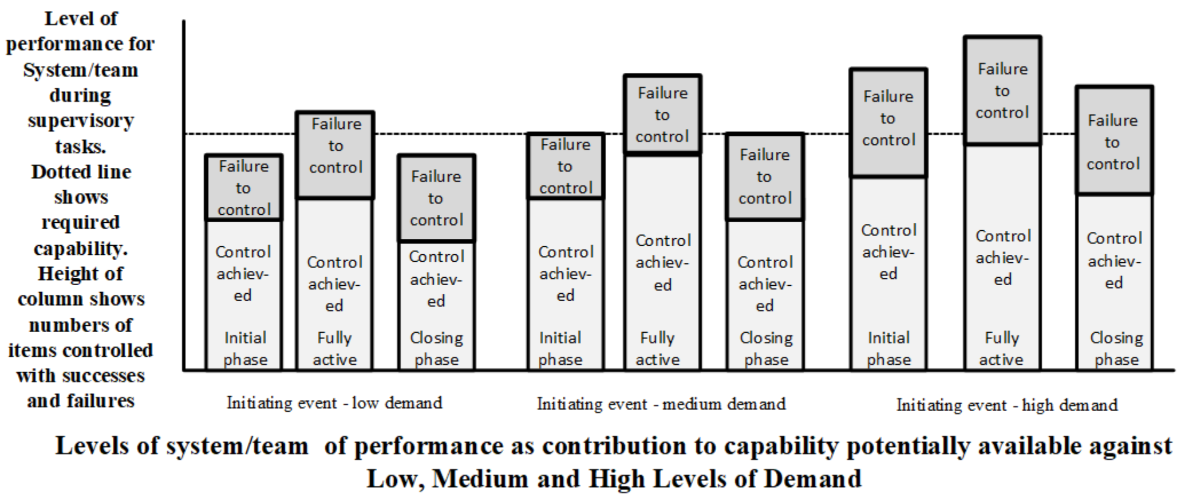


Figure 3: Capability available dependent on stage and level of demand

Conclusion

The precision and accuracy associated with the predictions given in Figure 3 may be lower than we might wish for. However this technique does offer a means of understanding better the consequences of design options when predicting capability. The model becomes available to the team - enabling debate. It has been used to show the sensitivity of design variables (e.g. allocation of function within the team) on capability.

In line with the definition of capability, it is possible to model initiating events and understand better the possible outcomes and areas of risk that may need to be addressed. Figure 3 showed that consequences of meeting three levels of demand.

This technique has been used in BAES to understand manning options for teams of up to three persons. This has enabled designs to be assessed and indications provided on bottlenecks associated with information flow. It has aided understanding of potentially important design issues while helping to understand which variables can safely be considered low risk.

References

Artman, H. (2000) Team Situation Assessment and Information Distribution. Ergonomics 43:8 pp 1111-1128.

Burns, C. M. (1999) Scanning Patterns with Ecological Displays When Abstraction Levels are Separated.

EMI Electronics (1979) Ergonomics for Future Submarine Systems. Ergonomics Laboratory Memo No 476

ISO 15288 (2013) System Life-Cycle Processes

Lindbom, H., Tehler, H., Eriksson, K. and Aven T. (2015) The capability concept – On how to define and describe capability in relation to risk, vulnerability and resilience. Reliability Engineering and System Safety Vol 135 pp 45 – 54.

Mitchell, P. J. and Cummings, M. L. (2005) Management of Multiple Dynamic Human Supervisory Control Tasks. The 10th International Command and Control Research and Technology Symposium

Sheridan, T. B. (1983) Supervisory Control of Remote Manipulators, Vehicles and Dynamic Processes; Experiments in Command and Display Aiding, ONR -Virginia.

Sheridan, T. B. and Hennessy, R. T. (1984) Research and Modelling of Supervisory Control Behaviour, Report of a Workshop, National Academy Press.

Tainsh, M. A. (1982) On Man-Computer Dialogues with Alpha-Numeric Status Displays for Naval Command Systems, Ergonomics, Vol 25(8) pp 683-703

Tainsh, M. A. (2018) User Systems Architectures - Two Studies in Design and Assessment. Applied Ergonomics Vol 68 pp 61- 71.

Using a systems thinking tool to identify work system interactions in healthcare

Emma Smith & Stacey Sadler

Nottingham University Hospital

SUMMARY

This paper advances the use of Systems Engineering Initiative for Patient Safety (SEIPS) to provide a visual representation of how work system factors interact with each other to shape processes and outcomes. Healthcare professionals identified that deficiencies in work system factors surrounding the person, task, tools and technology, environmental factors, organisational factors and external factors shaped undesirable outcomes for the patient, professionals or organisation around the discharge process. Improving work system factors may decrease the likelihood of negative outcomes for the patient, professionals or organisation.

KEYWORDS

Systems thinking, SEIPS, incident investigation

Introduction

Healthcare staff operate in complex systems, with many factors influencing the likelihood of errors. Adopting a systems approach is a term frequently associated with Human Factors (HF) and considers how the elements of the system interact with each other. HF models that are currently available aid our understanding of the dynamic interactions within the socio-technical system (Herrera and Woltjer 2008).

The Systems Engineering Initiative for Patient Safety (SEIPS) model 2.0 was developed primarily for use in the healthcare setting (Holden et al. 2013). It aimed to be a person centred sociotechnical framework, the work system produces work processes, which shape outcomes. Current diagrams utilised by SEIPS prevent key stakeholders to easily understand interactions between the work process and outcomes. Therefore, may not prioritise resource to areas where safety and efficiency can be significantly improved.

The aim of the Work System Interactions Map (WSIM) is to visually present the interactions between work system components that shape processes and lead to an outcome (positive or negative). The interaction map may increase the likelihood that recommendations can be targeted at system changes, increasing the likelihood of sustained safety improvements (Wheway. & Jun. 2021).

Method

Focus groups were conducted with healthcare professionals to discuss positives and challenges with the Trust’s discharge process. Broad questions were asked around positive and negative factors that impacted their job, challenges that most impacted on safety and identification of areas for improvement. Field notes were taken to capture the qualitative data. Thematic analysis was performed to identify broad themes that were discussed by participants either based on the number of times that participants mentioned the theme and those that had identified serious safety issues. The themes identified were categorised into the three broad areas Work System Factors, Processes and Outcomes. These were then mapped onto the WSIM and discussions with healthcare professionals and subject matter expertise were used to identify any interactions. This work focussed on work system factors that shaped undesirable outcomes to improve the discharge process.

Results

The WSIM highlighted that deficiencies in work system factors surrounding the person, task, tools and technology, environmental factors, organisational factors and external factors shaped undesirable outcomes for the patient, professionals or organisation around the discharge process.

Therefore, improving the whole system may have reduced the delay in patient discharge, decreased safety issues with eTTOs and improved patient experience of the discharge process. This also may have reduced blame on staff and improved staff job satisfaction. The WSIM further directs senior managers where effort should be made to address some of the deficiencies.

Discussion

The NHS Patient Safety Strategy strongly advocates the need for a systems approach that considers all relevant factors in the investigation of incidents and that the pursuit of safety should focus on strategies that maximise the frequency of things going right (NHS England and Improvement 2019). This paper presents a HF modelling technique based on SEIPS that aims to visually represent how the interaction of work system factors shape processes which can lead to an undesirable outcome.

More work is required to understand whether non-HF professionals can utilise this method and increase their knowledge and capability of applying a systems approach to understand the work system within healthcare, both proactively and reactively.

The development of the WSIM provides a visual representation to highlight interactions in a complex system easily. An overwhelming amount of support is available for a systems model that provides a visual representation of how work system factors interact whether in support of Serious Incident Investigation or transformational projects and it is hoped that this model starts this journey (Wheway & Jun 2021).

References

Herrera. I.A and Woltjer. R. (2008) Comparing a multi-linear (STEP) and systemic (FRAM) method for accident analysis . *Safety, Reliability and Risk Analysis: Theory, Methods and Applications*

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 factors framework for studying and improving the work of healthcare professionals and patients. *Ergonomics,* 56*(11),* 1669–1686.

Wheway. J.L & Jun. G.T. (2021). Adopting systems models for multiple incident analysis utility and usability. *International Journal for Quality in Health Care.* 33 *(4),* pp. 1-8.

NHS England and Improvement. (2019) The NHS Patient Safety Strategy: Safer culture, safer systems, safer patients. Report template - NHSI website (england.nhs.uk) Accessed 16th August 2022.

Envisaging regenerative futures through Good Work Design

Elise Crawford[1], Sara Pazell[1,2] & Nektarios Karanikas[3]

[1]CQUniversity Australia, [2]ViVA Health at Work, Australia, [3]Queensland University of Technology, Australia

ABSTRACT

As we move towards a fifth industrial revolution, concerns about the future of work are heightened. To answer the call for *work that we all want*, this paper extends the concept of Good Work Design (GWD) introduced by the Human Factors and Ergonomics Society in Australia in 2020. Following an overview of GWD, we present a list of respective features with the purpose to advance a human-centred design-led approach to workplace strategy that reconciles business success with worker health. Moreover, we argue that effective design practice should be regenerative, expanding capacity and capability for design throughout the organisation, while supporting sustainable futures. The goal of this paper is to stimulate ongoing debate, research, and practice in good work design.

KEYWORDS

Good Work Design; Ergonomics; Health; Wellbeing; Sustainability

Introduction

The ongoing globalisation of economies has transformed and continues to remodel the nature of work. As we move towards the fifth industrial revolution, concerns have been raised about emerging types of work. While earlier predictions viewed big-picture thinkers, collaborators, and people who can empathise with others as the twenty-first century skills (Pink, 2006), later positions suggest that today's global challenges require highly skilled workers with solid cognitive, interpersonal, and problem-solving abilities (Manyika et al., 2012). Indeed, new ways of working require new skill sets, but changes to working conditions are also taking a toll on worker health and wellbeing (Peters et al., 2022). On the one hand, there are calls for improved work conditions and opportunities to learn and grow at work, and, on the other hand, we experience an era of an increasing focus on strict compliance that stifles worker growth and development (Stein & Allcorn, 2020).

For instance, safety management systems based on a philosophy of achieving control by generating prescribed work procedures and commanding strict adherence to rules and regulations, neglect that the lack of worker input renders 'work as imagined' by managers incongruous to 'work as done' (Dekker, 2014). Admittedly, this approach may seem attractive to large and multi-national companies to ensure that businesses and workers comply with work health and safety regulations. However, imposed job designs have an infantilising impact on workers, which stifles motivation and personal development. Infantilisation has been found to lead to passivity and an overreliance on others and is limited to short-term gains (Alvesson & Spencer, 2017). Skills needed in crucial activities such as hazard identification and risk assessment are likely to diminish under these conditions, as with any personal motivation to innovate. A top-down approach opposes contemporary safety paradigms that explicitly encourage worker participation (Hollnagel et al., 2006) and more adaptive work processes (Provan et al., 2020).

Similarly, new ways of working, such as crowdsourcing and app-based work have led to the birth of web-based enterprises that support the gig economy but ushered in a new working class, labelled the ‘Precariat’ (Standing, 2016). The Precariat have less job security, earn less money, and have little to no health and safety provisions. These work conditions undermine the focus on sustainable development as per the UN’s (2023) agenda items to *ensure healthy lives and promote well-being for all at all ages* [Goal 3] *and promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all* [Goal 8]. Priorities for worker health and wellbeing are also evidenced by the growth in resources on this topic (e.g., US-OSG, 2022). Furthermore, negative work impacts have prompted industrial leaders to call for work design reform; for example, at the World Day for Safety and Health at Work 2019, respective discussions culminated in calls for a future of *work (that) we all want* (Mosier & Hiba 2019, p. 2).

To understand the impact of work design theory on management thinking and policy, Parker et al. (2017) conducted an extensive literature that revealed five distinct work design perspectives, namely Sociotechnical Systems Thinking and Autonomous Work Groups, Job Characteristics Theory, Job Demand-Control Model, Job Demand-Resources Model and Role theory. Moreover, their review indicated that work design is a key antecedent of most major focal areas of psychology and management, such as productivity, job satisfaction, wellbeing, absenteeism, presenteeism, organisational commitment, and creativity. Work design is found to play a mediating role among process and context variables, (e.g., leadership, downsizing, lean production, employment contracts) and business outcomes (e.g., productivity through job crafting) (Parker et al., 2017).

Another finding from the specific review was that traditional language used by academics (e.g., job design, work design, or job characteristics) was not used in daily practice. In industry, preferred terms may include job flexibility, collaboration, multidisciplinary teams, empowerment, future work, etc. and focus on addressing contemporary matters such as sustainability, globalisation, and ways to engage millennial staff (Parker et al., 2017). Additionally, the review of practical-oriented studies revealed mixed results, the sociological analysis of which suggested a rising trend towards more standardised work and lower decision-making autonomy in professional contexts.

Last, Parker at al. (2017) proposed a multilevel model of work design to bring the five work design perspectives together and address emergent issues arising at the individual level, social/system-level, and macro-level such as globalisation. Similarly, the Taylor’s review on modern working practices in the UK also called for responsible business that not only keeps pace with technology advancements and economic change, but also designs work that brings out the best in people and where work is founded on enduring principles of fairness (Taylor et al., 2017, p. 6). The need for good work has become a pressing matter of importance, heightening the demand for work design skills within the workplace that can aptly respond to current calls for work reform. The concept of Good Work Design (GWD) introduced by the Human Factors and Ergonomics Society of Australia (HFESA, 2020; Karanikas et al., 2021) responds to these calls for work reform.

Good Work Design: Overview

Design has been recognised worldwide, mainly through product design that improves life. In addition to Good Design® (2023) founded in Chicago in 1950, several countries have a Design Council or similar organisation such as the UK Design Council established in 1944 (UKDC, 2023) and Australian Design Council founded in 1958 (ADC, 2023). Albeit these and other organisations encourage quality designs through awards programs for physical items or structures, design is much more than that. For instance, the UK Design Council’s mission is “to make life better by design by working with people to create better processes, all of which lead to better performance.” (UKDC, 2020, p. 1). Also, the review by Parker et al. (2017) showed that although most countries have work design as a policy agenda item, government policies need to refocus from a mere emphasis on skill development to a greater emphasis on skill use within the workplace to achieve good work design.

This exact idea of *working with people to create designs that enhance their performance* is mirrored in and drives the concept of GWD: workers (including managers and employers) facilitated by human factors and ergonomic professionals or other specialists in work design, encouraged by management, supported by the organisation and educated by qualified experts to contribute to the design of their work, and continually build design literacy and capability within and across the organisation. Towards the end of 2019, this idea brought together a group of work design enthusiasts who formed a project committee within HFESA to craft a position on Good Work Design (HFESA, 2020). In alignment with the multidisciplinary nature of human factors and ergonomics, collectively, the committee represented ten discipline areas, namely the health sciences, social sciences, safety sciences, design science, psychology, engineering, legal services, education, human factors, and ergonomics, including representatives from Good Design Australia and the industry.

In principle, GWD is conceptualised as a fundamentally human-centred design-led approach that focuses on making good work available to all workers. 'All workers' extends from top executives to front-line workers, from maintenance staff to cleaners. Everyone in the organisation is there to do work, and hence, all are central to the success of the business. 'Good work' means that fundamental business objectives are realised while optimising human health and performance. The term 'good' denotes that there is no single endpoint of perfect work that can accommodate everything and everyone to the maximum, without trade-offs, especially within the reality of dynamically changing natural, socio-political, socio-technical, and organisational environments. 'Work design' does not follow a solid and rigid design process or outcome but it helps to ensure that the system of work is not a randomly and stochastically arranged and interacting set of agents. The term 'design' in GWD denotes the opportunity to continually co-conceive, co-create and redesign work in anticipation of and response to internal and external, systematic or random effects.

Achieving GWD involves three phases that are iteratively enacted and constantly adjusted as necessary: Discovery, Design, and Realisation (HFESA, 2020; Karanikas et al., 2021). In the Discovery phase, early engagement of individuals and teams is paramount. This includes those who drive design, those who co-design, subject matter experts, and those who may benefit from good work; often, these are the co-designers, but may include maintainers, and end users, like customers, or those within the supply chain. During the discovery stage, it is also necessary to study and comprehend the context, job, task, technology, equipment, and social interactions involved, so that problems can be defined, and opportunities noted. The Design phase involves collaboration, ideation, and facilitated solutions to problems, or the co-creation of opportunities for improvement. Activities may include simulations, prototype iterations, trials and reviews, the identification of trade-offs and negotiations. The Realisation phase refers to the tangible outcomes, deliverables developed as well as learning about their effectiveness, and optimisation levels. This phase seeks a balance between employee health and safety, productivity and other business outcomes. Figure 1 provides an overview of the GWD approach.

Good Work Design Features

Although the concept and elements of good work can be found in several publications, we advocate that GWD integrates, reconciles, and extends those. For instance, in the Australian context, literature has advocated for the design of good work to promote worker health (Kanse & Fruehn, 2022; SWA, 2020; AFOEM, 2011). Safe Work Australia (SWA) explains that good work means to manage risks and promote productivity and health (SWA, 2020), suggesting that by addressing worker health, productivity improves, and this supports the achievement of work objectives.

Although from the SWA's perspective this stands true and is supported by studies, in our view of GWD, healthy individuals and healthy businesses are equally important. In an elaborated vision of GWD, business objectives must be defined and met in tandem with worker health and wellbeing needs rather than the former objectives being a by-product of healthful work situations.

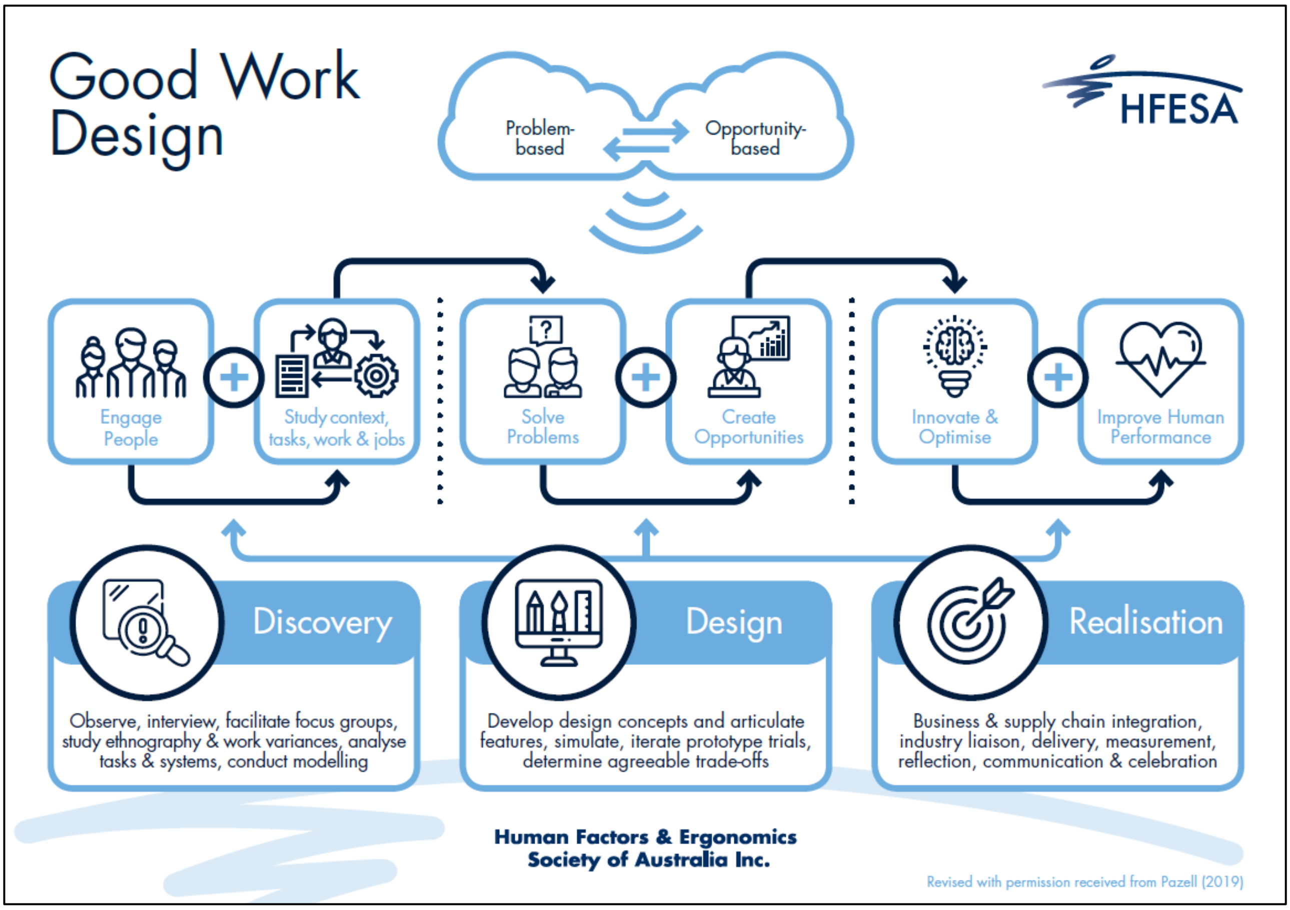


Figure 1: Good Work Design phases (HFESA, 2020)

Thus, while acknowledging that current published materials can be adequate if they match their targeted context and audience, we believe that there is space for a broader perspective that is more strongly oriented in design circles, compelled by human factors and ergonomics initiatives, and aligned with these practices. As such, and subject to ongoing discussions within and between academia and industry, we propose several provisional GWD features outlining what GWD is, what it acknowledges and appreciates, what GWD needs and does, and what it creates (Table 1).

We do not provide the list as an exhaustive checklist-type catalogue of features. Instead, we aim to stimulate further research and debate and, hopefully, poke those in positions of influence to apply +this holistic design framework to their organisational strategies. We appreciate that, in isolation, several of the features listed in the Table may mirror some ideas used in other literature or arise from a different orientation. However, contemplating, organising, and bringing all these features together under a unified, inclusive, design-oriented and discipline-agnostic GWD framework is innovative. Collectively, the GWD phases (Figure 1) and features (Table 1) represent a unified way to advance the agenda among different business units and their ontological framework or professional stance.

Table 1: Provisional list of Good Work Design features

Good Work Design	
IS	a framework for undertaking workplace (re)design
	human-centred
	propped by human factors and ergonomics approaches
	regenerative because it builds design capability (skills and resources) and capacity (ability to host and support design projects) throughout the organisation and the supply chain
APPRECIATES	the evolutionary and ecological aspects of variable human performance
	the "just right" balance of "joy work" and "work-work"
	'design-in-use' or the applications and spontaneous adaptations of work design in business
	the role of humans in highly automated systems
	that the design process is as important as the outcomes
	the positive emotional experiences associated with creation through design versus the fear associated with needs to contain all that can go wrong
NEEDS	facilitation by a work design strategist
	support by subject matter experts
	collaboration with 'conventional' designers
	resilience engineering strategies to inform and test work designs
	access to successful 'work arounds' or 'near rights' (versus 'near misses) to leverage on design-ready changes
	systems of transparent and defensible decision making in work governance
DOES	reconciles the varieties of human work
	designs for diversity
	more than 'consider business needs'; it realises business objectives in a competitive, pioneering, and sustainable manner while maintaining and promoting worker health and performance
	involves discovery, (iterative) design, and realisation of good work
	implements effective change management practice to test ideas, manage iterative trials, and launch progressively larger and more ambitious design campaigns throughout an organisation or cross-industry
	addresses safety-critical, material, unwanted, high-consequence and other types of unfavourable events (what is not wanted) but also focus on design for what is wanted for human performance across a spectrum of needs
	prospects new design opportunities on an ongoing basis
	tolerates a degree of fallibility to promote innovations
	focuses on storytelling to promote shared learning and tacit knowledge in a business
	enables cross-industry learnings and continual insights about a changing world of work
	celebrates design successes in a resounding way
CREATES	a visibly human-centred organisation
	unified business strategies among departments
	positive experiences of work, the effects of which extend beyond work
	a sense of coherence, meaning, and manageability to work
	a culture of innovation
	design that either works well or stands out because it is magnificent
	tacit knowledge about design to enable self-efficacy while building confidence to ideate, experiment, and innovate design-related change in supported or structured ways

The next steps

There are several opportunities to facilitate and support the implementation of the GWD approach. Although in this section we list the ones that we believe are most important currently, we remain confident that each reader, whether a scholar or an industry professional, can identify additional opportunities within their context. First, work design theories and studies need to become increasingly trans- and inter-disciplinary instead of viewing work from a mono-disciplinary or limited angle. Indicative necessary disciplines include, but are not limited to, design and safety sciences, human factors and ergonomics, operational engineering, business management and organisational psychology, appropriately complemented by experts from other disciplines depending on the work context.

Another opportunity regards industry-based projects that follow all three GWD phases from discovery to realisation and share best and poor practices. Instead of advertising only wins and great results, we must understand how compromises are made and what challenges arise. We need to gain honest and transparent insights through various channels (e.g., industry forums, conferences, publications, networking) as for example the successful and failed cases shared by authors from several countries and industries in two recent publications (Karanikas & Chatzimichailidou, 2020; Karanikas & Pazell, 2022). Implementing and testing the GWD phases across diverse work contexts will gradually build a crucial mass of knowledge to allow refinement of the GWD features and revisit its business value and merits.

To achieve the above, researchers need to design with the industry studies that go beyond cross-sectional surveys that collect perceptions or evaluate situations. We do not see GWD as another construct that represents, moderates, or mediates cause-effect relationships to be tested through hypotheses. GWD is about actioning its phases based on evidence- and practice-informed decisions, collecting data from the whole journey, and sharing all small and great struggles and wins. On this front, we must also improve the communication among practitioners, designers, researchers, and industry by presenting material in the language of the intended audience directed at contemporary concerns, so that audiences comprehend the relevance.

Conclusion

The approach to Good Work Design (GWD) through the extended concept presented in this paper responds to calls for better and fresh ways to design and manage work. The GWD features listed above illustrate our vision, but, most crucially, mean to advocate an informed, balanced, reconciled, and human-centred design-led approach to workplace strategy. We posit that this will enable business success and promote worker health and wellbeing. We promote GWD as a regenerative design practice that expands capacity and capability for design throughout the organisation and, thus, leads to sustainability in organisations. Nonetheless, we invite everyone to debate and challenge the content of this position paper and each other's views with the hope that the list of GWD features we have proposed will mature and advance GWD theory and practice.

References

ADC. (2023). Home. Australia Design Council. Available at: https://australiandesigncouncil.org/

AFOEM. (2011). Position Statement: Realising the health benefits of work. Australasian Faculty of Occupational and Environmental Medicine. Available at: www.healthbenefitsofwork.com.au

Alvesson, M., & Spencer, A. (2017). The stupidity paradox: The power and pitfalls of functional stupidity at work. London: Profile Books.

Dekker, S. W. A. (2014). The bureaucratization of safety. Safety Science, 70(1): 348-357. https://doi.org/10.1016/j.ssci.2014.07.015

Good Design (2023). GOOD DESIGN® the oldest and the most prestigious awards program, Available at: https://www.good-designawards.com/about.html
HFESA. (2020). Good Work Design. Human Factors and Ergonomics Society of Australia Available at: https://www.ergonomics.org.au/goodworkdesign/
Hollnagel, E., Woods, D. D., & Leveson, N. (2006). Resilience engineering: concepts and precepts. Aldershot, UK: Ashgate Publishing.
Kanse, L., & Fruhen, L. (2022) Work design. In Australian Institute of Health & Safety (AIHS). The core body of knowledge for generalist OHS professionals (2nd ed.). AIHS.
Karanikas, N., & Chatzimichailidou, M. M. (eds.) (2020). Safety insights: Success and failure stories of practitioners, Boca Raton: Routledge.
Karanikas, N., & Pazell, S. (eds.) (2022). Ergonomic insights: Successes and failures of work design, Boca Raton: CRC Press.
Karanikas, N., Pazell, S., Wright, A., & Crawford, E. (2021). The what, why and how of Good Work Design: The perspective of the Human Factors and Ergonomics Society of Australia. In Rebelo, Francisco (Ed.) Advances in Ergonomics in Design: Proceedings of the AHFE 2021 Virtual Conference on Ergonomics in Design. Springer, Cham, Switzerland: 904-911.
Manyika, J., Lund, S., Auguste, B., & Ramaswamy, S. (2012). Help wanted: The future of work in advanced economies. McKinsey Global Institute. March.
Mosier, K., & Hiba, J. C. (2019). The essential contribution of Human Factors/ Ergonomics to the future of work we want. International Labor Organization. Available at: https://www.ilo.org/global/topics/safety-and-health-at-work
Parker, S. K., Morgeson, F. P., & Johns, G. (2017). One Hundred Years of Work Design Research. Journal of Applied Psychology, 102(3): 403-420. https://doi.org/10.1037/apl0000106.
Peters, S. E., Dennerlein, J. T., Wagner, G. R., & Sorensen, G. (2022). Work and worker health in the post-pandemic world: a public health perspective. The Lancet Public Health, *7*(2), e188–e194. https://doi.org/10.1016/s2468-2667(21)00259-0
Pink, D. H. (2006). A whole new mind: why right-brainers will rule the future. New York City, New York: Riverhead Books.
Provan, D. J., Woods, D. D., Dekker, S. W. A., & Rae, A. J. (2020). Safety II professionals: How resilience engineering can transform safety practice. Reliability Engineering & System Safety, 195: 106740. https://doi.org/10.1016/j.ress.2019.106740
Safe Work Australia. (2020). Principles of good work design. Available at: https://www.safeworkaustralia.gov.au/
Standing, G. (2016). The Precariat: the new dangerous class. London: Bloomsbury Academic.
Stein, H. F., & Allcorn, S. (2020). The Psychodynamics of Toxic Organizations. 103–119. https://doi.org/10.4324/9781003009559-6
Taylor, M., Marsh, G., Nicol, D., & Broadbent, P. (2017). Good Work: The Taylor review of modern working practices. Available at: https://www.gov.uk/government/publications/good-work-the-taylor-review-of-modern-working-practices
UKDC (2021). Design perspectives: design skills. Design Council, UK. Available at: https://www.designcouncil.org.uk
UN (2023). The 17 goals. Sustainable Development. United Nations. Department of Economic and Social Affairs. Available at: https://sdgs.un.org/goals
US-OSG (2022). Workplace mental health & well-being. Office of the U.S. Surgeon General, Available at: https://www.hhs.gov/surgeongeneral/

Taking a systems approach to designing national safety policy

Tracey A Herlihey, Lauren Mosley & Matthew Fogarty

NHS England, United Kingdom

ABSTRACT

Any policy developed in a siloed manner and presented for implementation in a straightforward way is limited in its application in complex systems such as healthcare. In this article we describe a process for developing new patient safety policy by taking a user-centred approach and applying a system-based framework. The Patient Safety Incident Response Framework published in 2022 (NHS England, 2022), represents a complete redesign of how the NHS responds to patient safety incidents for the purpose of learning and improvement. The Framework will replace the current Serious Incident Framework (NHS England, 2015). Testing and revision were a formal part of the development cycle. The final version incorporates findings from an early adopter programme and independent evaluation and used SEIPS as a framework specifying the structure of a patient safety incident response system. We found the framework to be a useful tool for informing the revision of PSIRF; however, translating this work into policy form proved difficult and some nuance and direct links to SEIPS may have been lost.

KEYWORDS:

SEIPS, systems approach, patient safety, investigation, policy

Policy context

The current NHS approach to managing patient safety focuses on responding to patient safety incidents, as specified in the Serious Incident Framework (SIF) (NHS England, 2015). The SIF, and its predecessors, require organisations to investigate all patient safety incidents that are categorised as 'serious'. In 2023 a new framework, the Patient Safety Incident Response Framework (PSIRF) (NHS England, 2022), will replace the SIF in the English NHS. PSIRF represents a complete redesign of how the NHS manages patient safety.

Early development of PSIRF

The PSIRF has been in development for several years. In 2018 the National Patient Safety Team published 'The future of NHS patient safety investigation: engagement feedback' (NHS Improvement, 2018). This document described the findings from an engagement programme that aimed to seek the views from a wide range of stakeholders about how and when patient safety incidents should be investigated. This followed a previous 'investigation of investigations' led by the Team to provide a 'window on the system' (Vincent, 2004) of patient safety incident investigations in the English NHS.

Testing and revision: a formal part of the development cycle

Early adopters played a significant role in testing and revising the Introductory Framework (NHS England, 2020). The early adopter programme was integral in generating new knowledge about how the Introductory PSIRF could be implemented into practice. Mechanisms including monthly workshops and a formal independent evaluation of the programme were used to capture the new learning. The insight generated was in turn used to refine the framework.

Using SEIPS in policy design

The Systems Engineering Initiative for Patient Safety (SEIPS) provides a framework for understanding outcomes within complex socio-technical systems (Holden, et al., 2013). The framework has several uses as described in a recent paper published by Holden and Carayon (Holden & Carayon, SEIPS 101 and seven simple SEIPS tools, 2021). We used the framework prospectively to consider the design of an incident response system.

Building on the insight gathered during testing and revision we began by first considering the different outcomes across the various dimensions specified in the SEIPS 2.0 model including outcomes for patients, professionals, and organisations across dimensions of proximity and desirability. We then defined response processes or sequences of tasks to produce outcomes. Two broad processes were defined including capturing insight and transforming into improvement, and engagement and involvement of those affected by patient safety incidents.

Working through the SEIPS model we then went on to consider the various work system factors that may influence the defined process, which will in turn shape the outcomes we defined. Finally, we considered the potential for feedback loops from processes and outcomes to the work system and how these may present pathways to adapt the design of the system to ensure outcomes remain desirable.

Conclusion

The SEIPS framework has much potential in informing the design of improvement and innovation in healthcare. Here we have described how the framework was used in developing national patient safety policy. We found the framework to be a useful tool for informing the revision PSIRF and incorporating feedback from our early adopter programme and independent evaluation. The framework was helpful in highlighting the extent and importance of collaborative work required to produce the intended outcomes of a patient safety incident response system. Furthermore, the feedback loops within the framework prompted consideration of how an incident response system may adapt based on outcomes from learning response processes. This learning was incorporated into the new PSIRF. However, much reframing was needed when writing the final policy documents. The structure of the framework itself did not translate into a useful 'story' when writing policy, which meant some nuance and the direct links to systems engineering may have been lost.

References

Holden, R. J., & Carayon, P. (2021). SEIPS 101 and seven simple SEIPS tools. *BMJ Quality & Safety*, 901-910.

Holden, R. J., Carayon, P., Gurses, A. P., Hoonakker, P., Schoofs Hundt, A., 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*, 1669-1686.

NHS England. (2015). Serious Incident Framework.

NHS England. (2019). The NHS Long Term Plan.

NHS England. (2019). The NHS Patient Safety Strategy.

NHS England. (2020). Patient Safety Incident Response Framework 2020. *An introductory framework for implementation by nationally appointed early adopters*.

NHS England. (2022). Patient Safety Incident Response Framework.

NHS Improvement. (2018). The future of NHS patient safety investigation: engagement feedback.

Vincent, C. A. (2004). Analysis of clinical incidents: a window on the system not a search for root causes. *BMJ Quality & Safety*, 242-243.

Putting Ostomates at the Heart of Pouch Design

Anna McLister[1], Chloe Roberts[1] and A C B Medeiros[1]

[1] Kinneir Dufort Design Ltd. Bristol, BS1 5BU, UK

SUMMARY

Ostomy pouches are used daily by over 13,000 people in the UK each year, to collect effluent from their stomas. Although this Class I medical device has undergone a design revolution since the 1940s, ostomates' needs are still not being fully realised. Building upon knowledge and insights gained from interviewing and surveying ostomates, this paper will explore how applying key Human Factors considerations could help inform the future design of ostomy pouches and ultimately, improve the quality of life of ostomates.

KEYWORDS

Ostomates, Ostomy Pouches, Design, Human Factors, Healthcare

Introduction

Ostomates are often referred as people who have undergone surgery to create a stoma, which provides an opening at the surface of the abdomen to divert the flow of bowel or bladder effluent. This surgical intervention is conducted on 13,500 people in the UK each year and can be an acute or long term measure for a range of clinical conditions. This can vary from an illness, injury, or a problem with the digestive system, that can impact people of all ages from neonates to the elderly (Kettle 2019). A key product for ostomates are ostomy pouches which collect the stoma effluent. Simplistically, they consist of two key components; an adhesive baseplate to attach to the abdomen and a pouch to collect the effluent. Depending on the type of pouch (drainable or closed, one-piece or two piece system), ostomates may need to replace their pouch daily or up to once a week.

Since the development of commercially available rubber pouches in the 1940s (Lewis 1999), there has been a radical evolution in the design of ostomy pouches resulting in a variety of pouches available today. Despite these developments there has not been a major step change in pouch design since the 1980s, which means ostomates are still faced with product concerns and challenges because one pouch fit does not fit all. These challenges are evidenced in Quality of Life (QoL) studies which report that ostomates' QoL is impacted by skin problems due to failed skin barriers, pouching sizing systems, lack of access to support user groups and ostomy nurses (Maydick-Youngberg 2017). Therefore, this paper will explore the areas in which Human Factors considerations can be applied to help inform future design of ostomy pouches to help improve ostomates' QoL.

Applying Human Factors to pouch design

Considering that ostomy pouches are typically a Class I medical device within the EU, we wanted to explore through the lens of the Medicines and Healthcare products Regulatory Agency guidance (MHRA 2021) in relation to Human Factors considerations (users, use environment and device-user interface), and what this realistically means for ostomates.

Users

Users of ostomy pouches include a variety of people; from ostomates and lay caregivers to experienced and novice Healthcare Professionals (HCPs). With this range of users, comes various cognitive abilities (level of education, years of product experience, type of product training received and existing mental models), physical and sensory abilities (sex, age, their dexterity levels, their hearing abilities, their visual ability, and their co-morbidities) and user preference (pouch colour, pouch shape and size to match their daily actives) to take into consideration. Ultimately, how are ostomy manufacturers incorporating inclusive design principles to provide solutions for real people to help minimise their product frustrations and improve their QoL?

Use environment

Ostomy pouches can be applied and emptied by HCPs, caregivers and patients in various environments including a home setting, public setting, hospital, or community care setting. This of course introduces several variables: the room temperature (potentially impacting the adhesive properties of the baseplate), level of light (perception of design/instructions) and ambient noises (from home background noises to a hospital setting) which may impact how users interact and perceive the product, as well as their level of attention. How is the design of the pouch and instructions supporting users' needs when in various real-world environments?

Device-user interface

The user-interface of an ostomy pouch includes several touch points from the baseplate, baseplate liners, pouch material, filters, to a resealable tail (if drainable). With that comes a range of sensory experiences to consider; touch (how does the pouch material and shape feel against the body?) auditory (will the pouch materials crinkle and be easily heard by others?), sight (how does the location of their ostomy impact their ability to visually applying their pouch?) and smell (how effective are the filters in neutralising odour?). Overall, how is the design of the pouch interface supporting its ease of use and ostomate's pouch concerns?

Conclusion

It is evident that one type of ostomy pouch does not fit all ostomates. Echoing that of past research, more is needed to push the design and development of ostomy pouches into the 21st century to help alleviate ostomates' pouch concerns and unmet needs. Ultimately, we need to apply a holistic approach in the ostomy pouch design process to not only help improve product satisfaction and overall user experience, but more importantly the QoL of its users. This can only be achieved if ostomy pouch users are placed at the heart of Human Factors activities during the iterative design and development process.

References

Kettle, J. (2019). East of England NHS Collaborative Procurement Hub Integrated Care Team StoMap Programme Baseline Report. Accessed: Sep. 29, 2022. [Online]. Available: https://www.longtermplan.nhs.uk/wp-content/uploads/2019/08/nhs-long-term-plan-version-1.2.pdf.

Lewis, L. (1999). History and evolution of stomas and appliances. Stoma Care in the Community London. Nursing Times Books. pp.1-20.

Maydick-Youngberg D. (2017). A Descriptive Study to Explore the Effect of Peristomal Skin Complications on Quality of Life of Adults With a Permanent Ostomy. Ostomy Wound Manage.vol.63, no. 5, pp. 10-23.

MHRA, (2021). Guidance on applying human factors and usability engineering to medical devices including drug-device combination products in Great Britain. Accessed: Nov. 15, 2022. [Online]. Available: Guidance on applying human factors to medical devices - GOV.UK (www.gov.uk).

Identifying resilience: A system safety review of trauma and orthopaedic theatres

Victoria E. Wills[1], Andrew Seaton[1]

[1]Gloucestershire Royal Hospitals NHS Foundation Trust, UK

SUMMARY

A system safety review to assess the resilience in Trauma and Orthopaedic (T&O) theatres was conducted in response to a number of Never Events. The imminent publication of the Patient Safety Incident Response Framework (PSIRF) paved the way for an alternative to traditional serious incident investigation, proposing a systems-based approach and enabling subsequent improvements to be based on 'work as done', rather than 'work as imagined'. Analysis identified opportunities for interventions that built system resilience, which were developed and tested by front line staff as part of a Quality Improvement (QI) collaborative. The approach demonstrated a practical application of the integration of systems theory, patient safety, resilience engineering and quality improvement approaches.

KEYWORDS

Resilience, PSIRF, Quality Improvement, Systems, Work as Done, Safety II, SEIPS, CARe QI

Introduction

In April 2021, Gloucestershire Hospitals NHS Foundation Trust (GHNHSFT) reported two patient safety incidents that met never event criteria, taking the total to six, since March 2019, within the T&O specialty.

Never events are defined within the National Health Service (NHS) as, "Serious Incidents that are wholly preventable because guidance or safety recommendations that provide strong systematic barriers are available" (NHS Improvement, 2018, p.4). The traditional response to such events is the completion of a patient safety investigation that retrospectively identifies the factors that contributed to the undesired outcome, with the aim of making recommendations for improvements to prevent reoccurrence (NHS England, 2015). At GHNHSFT this approach had been followed for the preceding cases, however the repeated incidents indicated that the desired improvement was not being achieved.

With the imminent publication of PSIRF laying the foundations for the introduction of system-based analysis and improvement (NHS England, 2022), an alternative approach was proposed, which sought to analyse the system that had generated the undesired outcomes, rather than the undesired outcomes themselves.

The approach utilised the Systems Engineering Initiative for Patient Safety (SEIPS) (Carayon et al., 2006) and CARe QI (Anderson & Ross, 2020), to explore the system and identify opportunities to build system resilience. Staff were supported in applying a QI approach (Langley et al., 2009) to the findings from the systems analysis to develop and test interventions, based on the reality of 'work as done' (Hollnagel et al., 2015). Since the application of this approach, the median time between never events in theatres, has increased from 46 days to 224 days.

Method

National Requirements

The Serious Incident Framework (NHS England, 2015) describes the investigatory process required within the NHS when a patient safety incident meets the Never Event criteria. Obligations are met through the creation and sharing of an investigation report, detailing the factors that led to the unintended outcome. With PSIRF (NHS England, 2022), on the horizon, work is under way to introduce system-based analysis tools, however the Serious Incident Framework requirements still need to be met, during this transitionary period. A change in approach warranted a discussion with the (then) local Clinical Commissioning Group (CCG), where a system investigation was proposed in parallel with a related incident investigation. To ensure the existing requirements continued to be met, an investigation report detailing the circumstances of the incidents, previous findings and the intention to review the entire system responsible for the unwanted outcomes, was produced within the mandated 60-day timescale. To promote transparency and enable process governance, regular oversight meetings were established, to which the CCG were invited to monitor progress of this alternative approach.

System Analysis

With the scope of the review defined as 'procedures that involved implants, within trauma and orthopaedic theatres', high level process maps (Langley et al., 2009) were created with key staff, describing the intended process from patient identification through to patient recovery. These were created separately, for trauma and elective orthopaedic procedures. Due to covid-19 restrictions, they were constructed virtually using a google jam board (https:// jamboard.google.com/), the content of which was transferred to Microsoft Word for further review and amendment by the theatres staff. Although the process maps depicted 'work as imagined', they provided a sufficient outline of the process steps to enable the scope of the system review to be described and the next stage of the approach to be planned.

The systems analysis was conducted during a facilitated face to face workshop with approximately 40 multidisciplinary team (MDT) members from theatres, during which SEIPS (Carayon et al., 2006) was used. Staff were split into seven groups, one for each of the sections of the process that had been identified through process mapping. Individuals were allocated to a part of the process that they were familiar with, whilst being mindful of professional representation across the groups. After an introduction to the background and how to use SEIPS, the groups were tasked with identifying the system components, their interactions and their outcomes.

Following the SEIPS analysis, CARe QI (Anderson & Ross, 2020) was used to carry out observational studies of the theatres processes, with a focus on 'work as done' and with the aim of identifying indications of system resilience. Observers were introduced to the handbook, an overview of the project and the worksheets to be completed, through a virtual briefing over Microsoft Teams. Observations were scheduled across the same seven elements of the T&O processes that had been identified through process mapping and that had been subject to analysis using SEIPS. Observations were recorded on the worksheets provided by the CARe QI handbook and where necessary, to understand further what was being observed, questions were addressed to staff.

Theatres staff were notified during the morning team briefing when observations were to be conducted, as this was used as an opportunity to highlight the project that was under way and to provide reassurance around the purpose of the observations.

Completed worksheets were returned and reviewed with the aim of identifying evidence of the following resilience indicators within the observational descriptions:

- Anticipation
- Learning
- Adaptation
- Monitoring
- Responding
- Coordinating

Additionally, information on system outcomes and indications of misalignments in demand and capacity was noted. Resilience indicators were then used to construct a resilience narrative, which was used to identify improvement opportunities or areas that warranted further exploration.

Quality Improvement

To support the translation of improvement opportunities into improvement projects, a QI collaborative was established by the GHNHSFT Gloucestershire Safety & Quality Improvement Academy (GSQIA). Through this collaborative, 20 multidisciplinary staff from theatres undertook 5 QI projects to test and learn from potential interventions aimed at building system resilience in the areas identified through the analysis.

The collaborative was initiated by a day of virtual QI training, conducted over Microsoft Teams and based around the Model for Improvement (Langley et al., 2009). This included identifying a project aim, measures of improvement and change ideas, as well as showing how to test and assess change ideas using Plan-Do-Study- Act (PDSA) cycles. Figure 1, demonstrates the linkages between the tools and approaches.

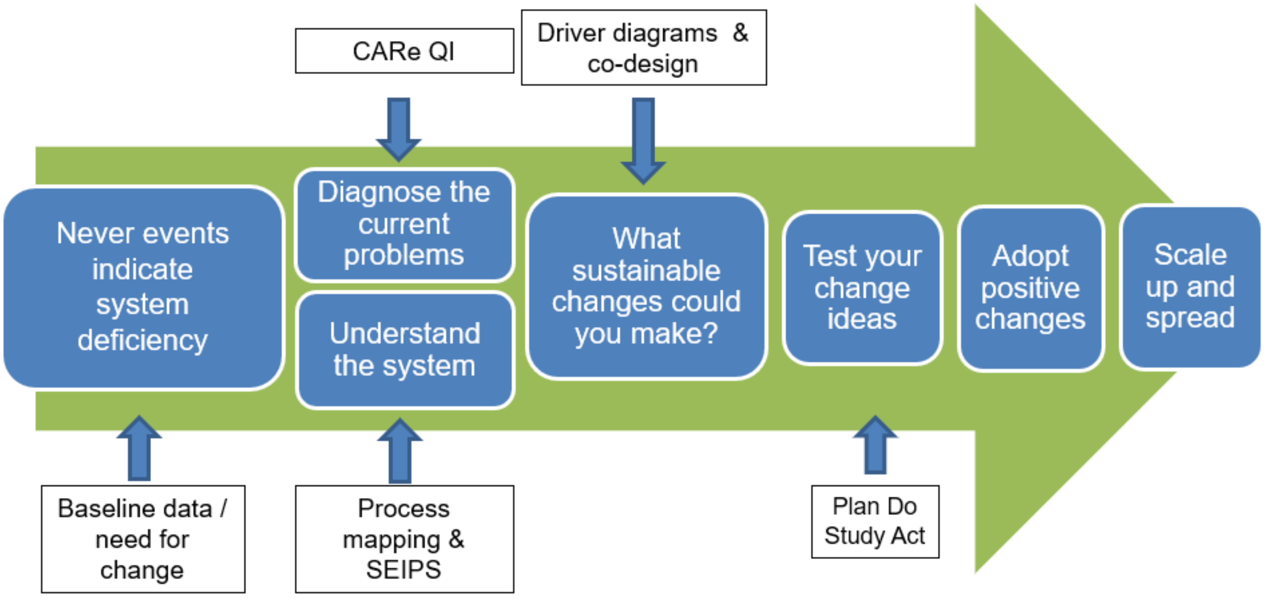


Figure 1: Linking system analysis with quality improvement

Teams were allocated a QI coach and a member of the Trust Human Factors Faculty to support them with their improvement work and over the following nine months, the teams worked to use QI methods to build system resilience. The work culminated in September 2022, with a celebration event where the teams presented their progress.

Although the majority of findings required an improvement approach, some areas identified by the systems analysis instead warranted management review or validation by further audit. Instead of being adressed through the QI collaborative, these were allocated and tracked through a governance meeting, established to oversee the progress of the project.

Results

From the high-level process maps, four process stages were identified for elective orthopaedic procedures and three for trauma, as shown in Table 1, below.

Table 1: Scope of System Review

Elective Orthopaedic	Trauma
1.Patient & implant identification, pre-assessment & listing.	1. Day before & day of procedure: Trauma list creation and amendment process.
2. Implant request, stock check, and preparation.	2. Day of procedure: Pre-list and pre-procedure implant checks.
3. Day of procedure: Implant collection and checking prior to patient arrival.	3. During procedure: In theatre implant checks.
4. Day of procedure: Implant checks prior to fit for trays/ sterile packaged components and loan items.	

The process stages were used to describe the scope of the SEIPS analysis and the observations. Each stage was allocated to an observer, with some observers covering more than 1 stage.

The resilience narratives constructed from the observations were compared with the SEIPS analysis, before being used to formulate the following recommendations identified in Table 2.

Table 2: Actions recommended by CARe QI

QI Project	Increase the successful completion of pre-assessment activities for elective orthopaedic cases.
	Improve the timely communication of necessary list changes within the two-week list 'lockdown' for elective T&O cases.
	Improving the storage of implants within theatres and the alignment of stock held with usage requirements.
	Improve the in-theatre checking process for implants.
	Improve the resilience of the 'golden patient' identification and notification, as part of the trauma list creation process.
Management Review	Review the demands on the role of the theatre coordinator.
	Review the impact of theatre utilisation requirements.
	Review capacity & demand of X-ray provision in theatres
	Risk Review of staffing and skill mix accounting for case demand & complexity.
Audit	Assess the availability and provision of sets for expected case load
	Assess the consistency of staff inclusion in the pre list WHO briefing.

As the system analysis had been instigated in response to never events within T&O theatres, data showing their reoccurrence within theatres was monitored. Figure 2 depicts the dates on which never events occurred and shows the number of days elapsed between events. At the time of writing (8th February 2023) it has been 422 days since the last Never Event in theatres. This is depicted by a dotted line on the chart as the data point marks merely the date used for measurement, rather than the occurrence of a never event. This represents an increase in the median days between theatres never events from 46 days to 224 days, since the start of the QI collaborative.

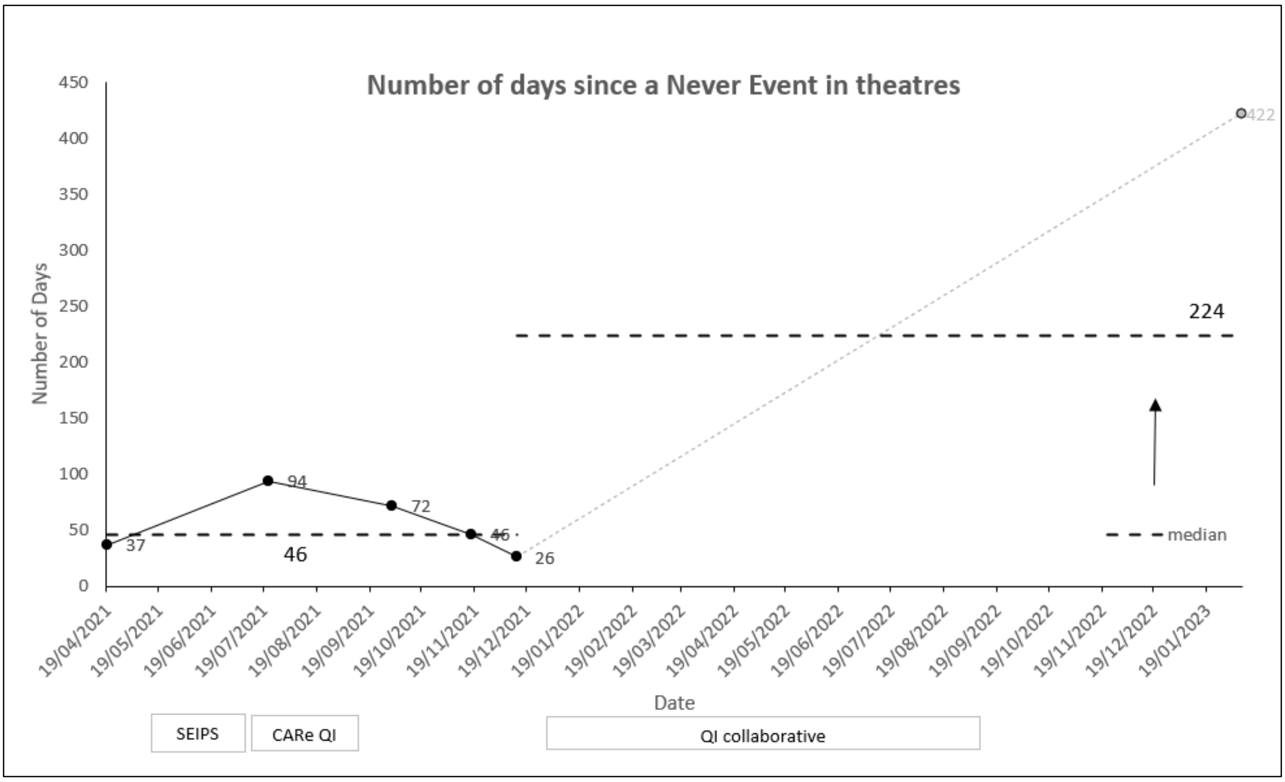


Figure 2: Time since last Never Event in theatres

Discussion

The approach described in this paper was instigated as a result of repeated Never Events within T&O theatres. Whilst thorough investigations had been previously carried out, the recommendations and actions had not been sufficient to prevent further occurrences. Developments within safety science recognise the limitations of a Safety I approach, (Anderson & Watt, 2020 & Hollnagel et al., 2015,) the basis for the traditional investigatory response to unwanted outcomes, when utilised within complex systems, such as healthcare. Similarly, the 'investigation – findings - recommendation – action' strategy, fails to take into consideration the body of evidence behind taking a quality improvement approach to develop and test changes to ensure effective and sustained improvement (Langley, 2009). The publication of PSIRF (NHS England, 2022), starts to correct these discrepancies and applying the approach as described, has presented an opportunity to test a systems-based approach that could be incorporated into a Patient Safety Incident Response Plan (PSIRP). The requirement for a patient safety incident investigation in response to a never event has, however, been retained within the PSIRF (NHS England, 2022).

The aim of the approach was a shift in focus from safety I and 'work as imagined' to safety II and 'work as done'. A Safety I approach could not be avoided in its entirety, as the review was triggered by the never events and the SI framework still required an investigation report. The system safety review, however, took a safety II approach and sought opportunities to build system resilience, changing the focus from '*ensuring that 'as few things as possible go wrong' to ensuring that 'as many things as possible go right''*. (Hollnagel et al., 2005, p.4). Utilising the CARe QI handbook, enabled the adaptations and variations within the system to be witnessed, whilst SEIPS supported the collection of a diversity of views. These in combination helped to capture the complexity and the reality of 'work as done', whilst recognising the value of staff engagement through involvement in the analysis and improvement processes.

The change in approach supported a wider scope of review, leading to the identification of latent factors and improvement needs in areas that had not been identified through the previous traditional Safety I investigation processes undertaken by GHNHSFT. This supports the findings from those such as Anderson & Watt (2020), Hollnagel et al. (2015) & Wigg et al. (2020) who have highlighted the limitations of the investigation process.

Limitations

The application of the approach was affected by varying restrictions associated with the Covid-19 pandemic, which impacted on staff availability due to sickness and the ability to arrange staff gatherings, due to restrictions in group size and proximity. This required the approach to be modified in the following ways:

- Limitations in the number of staff allowed to gather in indoor spaces resulted in the SEIPS workshop being hosted across multiple rooms, limiting the team interactions and knowledge sharing.
- QI teaching was conducted virtually over MS Teams, replacing the preferred format and length of session offered by a face-to-face workshop.

Access to front line staff was an essential component of the approach but also the greatest challenge. Shortages led to the process taking longer than anticipated and limited the number of observations that would have otherwise been conducted to get greater system representation. This was also the greatest challenge for the teams working on the quality improvement projects as their time was limited by the continual need to staff theatre lists, due to limited staff availability.

Whilst the use of CARe QI enabled the reality of the work system to be observed, it was not possible to observe all shifts, days of the week, teams at work or variation in process. The observations therefore represented a sample of the system at work and so may not have been representative of all permutations, or captured all factors that limited system resilience. Similarly, any issues that were observed, may have been over represented due to the limited sample size of the observations conducted. This risk was mitigated to some extent by utilising two system analysis tools, so that their outputs could be compared. This comparison did not highlight any missed opportunities or anomalies in the findings.

Completing this system safety review also required a trade-off between enhanced scope and complexity and the time required. The length of time that that it took for this approach to be developed, agreed and implemented was substantially longer than the prescribed 60-day limit for an SI investigation. Whilst the investigation report produced in parallel to meet the SI obligations was completed within the 60 days' timescale, the recommendations were linked into the system safety review. Being permitted the time to undertake the approach without the development of the standard action plan, required a degree of negotiation due to historical expectations of the investigation process and its outputs.

Further practical considerations included the necessity to split the system to be reviewed into sections identified through the use of process mapping. Whilst this precluded an entirely holistic approach to the analysis, it was a necessary response to the complexity of the system and to enable the logistics of the approach to be managed. These logistics were subsequently coordinated by the Trust safety department, in conjunction with a member of theatres staff. Having an individual point of contact within the theatres team proved to be an essential component in the coordination and delivery of the approach, as they were able to apply their knowledge of the teams and the processes to the planning and implementation, such that it had the best chance of success and theatre team involvement.

Each of the individual QI projects were tracked through their own identified measures, whilst the overall impact of the approach considered the recurrence of never events. Whilst this data indicates a substantial increase in time since the last never event, it is not possible to attribute this to the system safety review alone, due to the lack of a control group for comparison and the inability to control the myriad of variables within the operational environment.

Conclusion

The approach required advance planning, staff involvement and a considerable amount of time and coordination, in exchange for a much broader, system focussed review, based on work as done, rather than work as imagined. The advent of PSIRF can support the use of such approaches more regularly in the future, however organisationally this will require a shift in expected timescales, staff involvement and outputs. The benefit of the system analysis is multi-faceted – from the intelligence that the process unearths about complex systems to its ability to highlight the adaptability of staff through the differentiation between work as imagined and work as done. Additionally, its ability to involve and engage staff in the diagnostic and improvement process should not be underestimated.

With PSIRF now published, the learning from this practical application of systems analysis, resilience engineering and quality improvement, can be used to inform the development of PSIRPs within the NHS, providing an opportunity in the future to focus on '*ensuring that 'as many things as possible go right''.* (Hollnagel et al., 2005, p.4).

References

Anderson, J. E. & Ross, A. (2020). CAREe QI: A handbook for improving quality through resilient systems. https://resiliencecentre.org.uk/care-qi-handbook/

Anderson, J. E. & Watt, A. J. (2020). Using Safety-II and resilient healthcare principles to learn from Never Events. *International Journal for Quality in Health Care, 00*(00), 1–8. doi: 10.1093/intqhc/mzaa009

Carayon, P., Schoofs Hundt, A., Karsh, B. T., Gurses, A. P., Alvarado, C. J., Smith, M., Flatley Brennan, P. (2006). Work system design for patient safety: The SEIPS model. *BMJ Quality & Safety, 15*(1), i50-i58. https://doi.org/10.1136/qshc.2005.015842

Hollnagel, E., Wears, R. L. & Braithwaite, J. (2015). *From Safety-I to Safety-II* [White Paper]. University of Southern Denmark, University of Florida, USA, and Macquarie University, Australia. https://www.england.nhs.uk/signuptosafety/wp-content/uploads/sites/16/2015/10/safety-1-safety-2-whte-papr.pdf

Langley, G. J., Moen, R. D., Nolan, K. M., Nolan, T. W., Norman, C. L. & Provost, L. P. (2009). *The improvement guide. A practical approach to enhancing organizational performance* (2nd ed). Jossey-bass.

NHS England. (2015). *The Serious Incident Framework*. https://www.england.nhs.uk/wp-content/uploads/2020/08/serious-incidnt-framwrk.pdf

NHS England. (2022). *Patient Safety Incident Response Framework.* https://www.england.nhs.uk/wp-content/uploads/2022/08/B1465-1.-PSIRF-v1-FINAL.pdf

NHS Improvement. (2018). *Never Events policy and framework.* https://www.england.nhs.uk/wp-content/uploads/2020/11/Revised-Never-Events-policy-and-framework-FINAL.pdf

Wiig, S., Braithwaite, J. & Clay-Williams, R. (2020). It's time to step it up. Why safety investigations in healthcare should look more to safety science. *International Journal for Quality in Health Care, 32*(4), 281–284. https://doi.org/10.1093/intqhc/mzaa013

Trap and drag incidents on the London Underground – the role played by passenger mental models

Jasmine Bayliss[1, 2], Patrick Waterson[2] & Mark Young[1,2]

[1]Human Factors and Complex Systems Group, Loughborough University, UK, [2]Rail Safety and Standards Board (RSSB), London, UK, [3]Rail Accident Investigation Branch (RAIB), Farnborough, UK.

ABSTRACT

A trap and drag accident occurs when a passenger gets part of their body, or an object trapped between the train doors and gets dragged along the platform as the train departs causing various consequences of varying severity (Roels, 2018). The primary aim of this study was to understand why trap and drag accidents occur, how they come about and, establish how they can be prevented. A secondary aim was to understand how frequently traps occur and go unreported, and to understand if passengers are aware of the risks at the Platform-Train-Interface (PTI). We found that the majority of participants expect the train doors to work in the same way as lift and automatic shop doors and believe that the doors would automatically reopen if there was an obstruction. This study has also found that 40% of in-person survey participants and 27% of online participants have experienced becoming trapped. 6% of in-person survey participants have experienced being trapped and dragged. None of these incidents were reported. This study shows that passenger mental models of train doors are confused and inconsistent. Passenger's experiences do not correlate with their expectations of how the doors work, their mental models are incomplete, and this can lead to passengers taking risks when boarding or alighting.

KEYWORDS

Rail passenger safety, platform-train-interface, trap and drag incidents

Introduction

From April 2020 to March 2021 there were 2,042 injuries to passengers in stations or on trains (Office of Rail and Road, 2021). 13.5% of severe accidents occurred at the platform edge while passengers were boarding or alighting the train, including trap and drag accidents (RSSB, 2021). A trap and drag accident occurs when a passenger gets part of their body, or an object trapped between the train doors and gets dragged along the platform as the train departs causing various consequences of varying severity (Roels, 2018). The primary aim of this study was to understand why trap and drag accidents occur, how they come about and, establish how they can be prevented. A secondary aim was to understand how frequently traps occur and go unreported, and to understand if passengers are aware of the risks at the Platform-Train-Interface (PTI).

Methods of study

The study involved the use of a mixed set of methods. An in-person survey was carried out on 53 participants; these surveys took place at four different stations. An online survey was also conducted with 102 participants. 21 RAIB reports on trap and drag accidents were analysed and the causal factors for each recorded. In addition, 5 interviews with industry experts have also been conducted as part of this study.

Findings

We found that the majority of participants expect the train doors to work in the same way as lift and automatic shop doors and believe that the doors would automatically reopen if there was an obstruction. This study has also found that 40% of in-person survey participants and 27% of online participants have experienced becoming trapped. 6% of in-person survey participants have experienced being trapped and dragged. None of these incidents were reported. This study shows that passenger mental models of train doors are confused and inconsistent. Passenger's experiences do not correlate with their expectations of how the doors work, their mental models are incomplete, and this can lead to passengers taking risks when boarding or alighting. It was found that although the majority of participants are aware of the possibility of traps and the risks of obstructing the doors their motivations to board are stronger than their fear of entrapment or injury. The study also suggests that traps and traps and drags happen more frequently than previous research shows.

Discussion

The data shows that passenger expectations of what happens when there is an obstruction differs from actual events. Most participants expect the doors to reopen automatically when it detects an obstruction and do not understand the potential severity of their actions. We also show that the traps and traps and drags occur more frequently than previous research shows. The most popular response to passenger's expectations of what would happen if there was an obstruction of the doors, was that the doors automatically reopen. This was shown by 68% of participants from the in-person survey and 47% from the online survey. Other expectations given for what would happen if there was an obstruction included, the doors reopening depending on the size of trapped object, an alert being sent to the driver, and the emergency button being pressed. 4% of responses in the in-person survey and 17% in the online survey thought that injuries, drags, or the doors remaining shut were possible outcomes of entrapment. This shows that most participants do not expect serious consequences following entrapment, this belief can lead to passengers carrying out risky behaviours which are more likely to lead to trap and drag accidents because they do not expect them to be a possibility.

From the data collected two mental models of how passengers expect the train doors to work have were created. The first model is for those participants who believe that the doors will automatically reopen if there is an obstruction. The second model is for those who do not expect the doors to open automatically and instead expect there to be a mechanical element to the doors which either sends a signal alerting the driver of the obstruction or causes the doors to stay closed. The mental models created are incomplete and confused models if looked at on an individual basis, many who believe that the doors work in the same ways as lifts have had experiences becoming trapped and have had to push the doors open themselves to become free, others who expect the doors to work this way understand that fingers and coats can become trapped in the doors and be undetected. Passengers using the second model expected the doors to automatically reopen if rucksacks, arms, or fingers became trapped but when asked separately what happens when something gets trapped thinks a signal is sent to the driver or the door button must be pushed for the doors to reopen.

Conclusions and future work

In conclusion this study shows that passenger mental models of train doors are confused and inconsistent. Passenger's experiences do not correlate with their expectations of how the doors work, their mental models are incomplete, and this can lead to passengers taking risks when boarding or alighting. The study also suggests that traps and traps and drags happen more frequently than previous research shows. The majority of participants expect the doors to automatically reopen if there is an obstruction just like lift doors. These findings are consistent between the two surveys as well as with the research carried out with the RSSB (2017). This misconception can lead to passengers believing that there is little risk in boarding late or obstructing the doors which can lead to an increased risk of the passenger becoming trapped and potentially dragged. The paper concludes with a set of recommendations for future work, as well as implications for the design of interventions to improve passenger safety a the PTI (e.g., the design of warnings and public messaging campaigns).

References

Office of Rail and Road, 2021. Rail Safety 2020-21. [online] Office of Rail and Road. Available at: <https://dataportal.orr.gov.uk/media/1999/rail-safety-2020-2021.pdf> [Accessed 14/11/2022].

Roels, R., 2018. Rail passenger behaviour invites ‘trap and drag’ incidents. [online] DNV. Available at: <https://www.dnv.com/article/rail-passenger-behaviour-invites-trap-and-drag-incidents-199979> [Accessed 14/11/2022].

RSSB, 2017. T1102-Optimising door closure arrangements to improve boarding and alighting. London: Rail Safety and Standards Board.

Recreational Boating Safety: A Systems Analysis of the Causal Factors Contributing to Accidents

Helen Wordsworth[1], Patrick Waterson[1] & Will Tutton[2]

[1]Human Factors and Complex Systems Group, Loughborough University, UK, [2]Marine Accident Investigation Branch (MAIB), Southampton, UK.

ABSTRACT

Recreational boating has become an extremely popular past-time in the UK, particularly since COVID-19, with boat equipment sales up by 25%, compared with pre-pandemic levels. Wilson (2022) predicts a further 11% growth in sales in 2021-2022 and the market is currently exceeding pre-pandemic trading levels. We used a set of 12 Accimaps to analyse of contributory factors leading to recreational boating accidents documented by the MAIB. The data from the Accimaps was then used to scope a set of questions which formed the basis of two surveys. One survey was for members of the public involved in recreational boating and the other for professional individuals involved in the recreational boating community. Key findings were that lack of training, knowledge and preparation were seen as key reasons leading to unsafe boating situations, in addition to recklessness. Most participants wore lifejackets whilst boating and approximately 50% thought alcohol consumption was acceptable at some point during a boating trip. Boat/sailing clubs were seen to reinforce good safety culture but outside of clubs was less positive. Some recommendations were developed following the study. It would be beneficial to introduce a mandatory qualification prior to the purchase of any motorised vessel, such as that employed in Australia or similar to the International Certificate for Operators of Pleasure Craft (ICC).It may be worthwhile developing more interactive, nationally consistent signage at popular launch locations. A final recommendation would be to place some responsibility on manufacturers.

KEYWORDS

Accidents at sea; perceptions of safety; recreational boating

---

Introduction

Recreational boating has become an extremely popular past-time in the UK, particularly since COVID-19, with boat equipment sales up by 25%, compared with pre-pandemic levels. Wilson (2022) predicts a further 11% growth in sales in 2021-2022 and the market is currently exceeding pre-pandemic trading levels. The South East and South West have the highest distribution of boating activity (RYA, 2022). The increase in recreational vessel/craft (RC) use has led to an increase in accidents (MAIB, 2022), with total UK marine accidents consistently increasing since 2019, from 1090 accidents to above pre-pandemic levels, at 1530 accidents in 2021. National Water Safety, NWS (2022) documented that although accidental water-related fatalities decreased slightly in 2021, the three-year average was above that of previous years. 83% of casualties were male and 55% were due to recreational activity (NWS, 2022). The research aimed to investigate the culture and attitudes towards recreational boating safety in the UK; to identify the specific underlying conditions and active failures that lead to recreational boating accidents; and to identify what kind of recommendations may positively influence recreational boating safety culture in the future.

Methods of study

We used a set of 12 Accimaps to analyse of contributory factors leading to recreational boating accidents documented by the MAIB. The data from the Accimaps was then used to scope a set of questions which formed the basis of two surveys. One survey was for members of the public involved in recreational boating and the other for professional individuals involved in the recreational boating community, such as enforcement agencies, sales and emergency services personnel. Total sample size was 92 participants, of which 28 were professionals and 64 were members of the public. Surveys were distributed online and in-person. Results were analysed using descriptive statistics and qualitative thematic analysis.

Findings

Key findings were that lack of training, knowledge and preparation were seen as key reasons leading to unsafe boating situations, in addition to recklessness. Most participants wore lifejackets whilst boating and approximately 50% thought alcohol consumption was acceptable at some point during a boating trip. Boat/sailing clubs were seen to reinforce good safety culture but outside of clubs was less positive. Legislation and regulation surrounding use of recreational vessels in the UK is currently under-developed and most survey participants viewed mandatory training as an appropriate intervention to improve safety. Both survey groups stated that lack of knowledge, awareness for weather and tides and reckless behaviour contributed to accidents. Other people's behaviour was also highlighted by the public, with many specifically mentioning PWCs. Lack of preparation and alcohol was also a theme for both groups. Non-use of killcords was given a much higher focus from the general public than professionals, together with lack of safety equipment. However, complacency, whilst a strong theme from the professional group, wasn't mentioned by the public.

In the UK, the research focus has been on recreational boating fatalities and direct causes, such as drowning. Little is published on contributing factors, but the U.S. Coastguard publish these annually. Due to the similarity in Western cultures, these statistics are arguably comparable to accidents in the UK. Between 2020-2021 in the USA, the key contributing factors were inexperience, poor lookout, speeding and vessel maintenance/failure (Arguin, 2021). AcciMap themes included lack of knowledge/experience, training and qualifications, which supports evidence by Miller and Pikora (2007) and U.S. coastguard data.

Discussion, conclusions and future work

A range of attitudes have been revealed surrounding recreational boating safety. In conclusion, it appears that the majority of boat users believe motorised vessels to be most problematic, particularly personal watercrafts (PWCs), which was reinforced by the AcciMap analysis of MAIB reports. Furthermore, lack of knowledge and training was identified as the predominant theme that jeopardises safety at sea. Problems with enforcement were recognised as the biggest potential factor hindering progress to safety improvements, but participants agreed that a type of licensing may be the best way to make recreational boating safer. Some recommendations were developed following the analysis. It would be beneficial to introduce a mandatory qualification prior to the purchase of any motorised vessel, such as that employed in Australia (Virk and Pikora, 2010) or similar to the International Certificate for Operators of Pleasure Craft (ICC). Although safety advertising was generally dismissed, it may be worthwhile developing more interactive, nationally consistent signage at popular launch locations. For example, digital weather reports, which could be updated daily; and a traffic light system for clearly informing individuals of conditions. A final recommendation would be to place some responsibility on manufacturers. One professional participant owned a PWC sales business and with every vessel sold, an RYA personal watercraft course was included. Although the onus was then on the individual to book the course, this would undoubtedly increase uptake of course participation, as it would negate cost, a primary theme in non-attendance.

References

Arguin, W. R., 2021. 2020 Recreational Boating Statistics [online]. Washington DC, USA: U.S. Department of Homeland Security. Available at: https://www.iims.org.uk/wp-content/uploads/2021/07/US-Recreational-Boating-Statistics-2020.pdf [Accessed: 21/11/2022]

National Water Safety (NWS), 2022. WAID UK 2021 Summary for the NWSF. Available at: https://www.nationalwatersafety.org.uk/media/1315/waid-uk-2021-summary-11-final.pdf [Accessed: 21/11/2022]

Marine Accident Investigation Branch (MAIB), 2022. Marine Accident Recommendations and Statistics. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1080967/MAIBAnnualReport2021.pdf [Accessed: 21/11/2022]

Royal Yachting Association (RYA), 2022. UK Coastal Atlas of Recreational Boating. Available at: https://www.rya.org.uk/knowledge/planning-licensing/uk-coastal-atlas-of-recreational-boating [Accessed: 21/11/2022]

Virk, A. and Pikora, T., 2010. The Recreational Skipper's Ticket and its influence on boater behaviour. International Journal of Aquatic Research and Education, vol. 4, pp. 175-185.

Women Are Not Small Men

Laird Evans[1], Robert Pringle[2] & Eluned Lewis[2]

[1]MOD Defence Equipment & Support, [2]QinetiQ

ABSTRACT

Human Factors Engineers need accurate anthropometric data to design military equipment that is safe, comfortable and enables performance under extreme operational conditions and in the most severe environments. MOD acknowledges that its current anthropometry dataset is becoming increasingly unrepresentative of today's Armed Forces personnel, particularly women and minority ethnic groups. To address this issue, MOD has launched a new, comprehensive anthropometry survey. This paper describes the requirements underpinning this survey and the planned solution.

KEYWORDS

Anthropometry, Anthropometric, Survey, Body Armour

Introduction

Human Factors Engineers need accurate anthropometry data to design equipment that is fit for purpose – i.e. equipment that is safe and comfortable, and which yields optimal and acceptable performance. This is true regardless of the domain and application being considered. It is absolutely true in the Defence domain, when designing military systems, equipment and clothing for use under extreme operational conditions and in the most severe environments. Yet Human Factors Engineers in the Defence sector are using data that are acknowledged by the Ministry of Defence (MOD) to be not fully representative of today's Armed Forces personnel. This is particularly so for women and minority ethnic groups. The following quotation is taken from the House of Commons Defence Committee, Second Report of Session 2021-22:

"We find it extraordinary that uniforms and equipment are still a problem across all services. Thousands of female Service personnel, already facing the dangers of military duty, are at greater risk of harm due to basic failures in their uniform and equipment, which can have consequences for their combat effectiveness and health. Fixing these problems is one of the simplest ways that the Forces can demonstrate they value their servicewomen."[1]

It is clear from this quotation that the safety, health, comfort and performance of service women may be being compromised, but this is true also for minority ethnic groups, small men and, arguably, for all users of military equipment. (The term "equipment" is used in a general sense to mean systems, sub-systems, individual components, clothing, Personal Protective Equipment (PPE) and even whole platforms.)

The last comprehensive anthropometry survey of UK Armed Forces personnel was conducted in 2006-07 (Pringle et al, 2011). The data gathered at that time are freely available in the MOD's Human Factors Integration (HFI) Technical Guide for Anthropometry: People Size (Cummings,

[1] Protecting those who protect us: Women in the Armed Forces from Recruitment to Civilian Life - Defence Committee - House of Commons (parliament.uk).

2022). It is principally these data that Human Factors Engineers in MOD and Industry use when designing military equipment.

To address the lack of wholly representative and up-to-date data, the MOD's Defence Ordnance Safety Group (DOSG) placed a contract with QinetiQ to conduct a new, comprehensive, tri-service anthropometry survey. This paper describes the requirements placed on QinetiQ and the solution proposed in response.

Requirements

Two workshops were conducted in 2021 to capture requirements for anthropometry data from MOD stakeholders. These workshops were organised and facilitated by Human Factors specialists in the MOD's Defence Equipment & Support (DE&S) Human Factors Integration (HFI) team. These specialists were contracted to support DOSG during the early phase of requirements gathering and remain engaged to support the conduct of the anthropometry survey.

Both workshops attracted representation from many MOD establishments and teams. Following these events, key stakeholders emerged and their requirements for data were prioritised. At the same time, principal points of contact for the three services – Army, Royal Air Force (RAF) and Royal Navy (RN) – were identified. In addition, the survey has attracted support from the highest levels in the MOD and was endorsed by Lieutenant General James Swift, Chief of Defence People (CDP); Lieutenant General Ivan Jones, Commander Field Army (CFA); Rear Admiral Jude Terry, Director Personnel and Training (RN); and Air Marshal Sir Gerry Mayhew, Deputy Commander Operations (RAF).

Body Armour Requirements

Of particular interest during the requirements gathering phase were the bodily dimensions to be measured. The 2006-07 survey had measured 92 dimensions from 2470 personnel. It was ultimately determined that all of the dimensions recorded previously should also be measured in the new survey. In addition, a significant number of new measurements are required to support the development of new body armour. It is widely acknowledged that *"women are not small men"* (Lewis, 2020) and this is never more critical than when designing body armour that will not only save lives, but which has the potential to reduce the incidence of musculoskeletal injuries, the most common cause of medical downgrading and medical discharge in both Service men and women (MOD, 2016).

The development of body armour is a complex undertaking, involving significant engineering trade-offs. Not only must the armour protect the wearer, but it must do so in a manner that reduces restriction of movement to a minimum and does not incur unacceptable thermal or physical discomfort. Most importantly, the armour must offer protection to the critical areas of the body. A study by Breeze *et al*. (2016) identified protection of the heart, great vessels[2], liver and spleen to be of paramount importance if death or significant long-term morbidity is to be avoided. Therefore, to enable new body armour to be designed and optimised to protect these organs, 53 new dimensions are required to be measured. Examples of two of these critical dimensions are illustrated in Figure 1. These are:

A. Suprasternal notch to tenth rib
B. Suprasternal notch to iliac crest

[2] The large arteries and veins directly connected with the heart are termed the great vessels, consisting of the inferior vena cava, superior vena cava, pulmonary arteries, pulmonary veins, and root of the aorta.

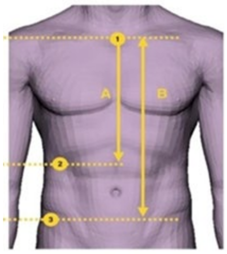


Figure 1: Critical Dimensions (A & B) from Breeze *et al*, 2016

In addition to the measurement of new external dimensions, the development of body armour to protect the anticipated user population – male and female, across all services – requires the measurement of the sizes and positions of the critical internal organs. A separate but related study, also initiated and funded by DOSG, was launched to acquire the necessary data. This study is being led by the Defence Science and Technology Laboratory (Dstl) and will utilise a Magnetic Resonance Imaging (MRI) scanner based at the University of Nottingham. Therefore, an additional requirement is for QinetiQ to collaborate with Dstl to ensure that the two studies yield complementary data, such that the combined dataset meets the requirements for new body armour.

Data Analysis Tool

It is important to note that the requirements placed on QinetiQ do not include a requirement to develop a new anthropometry data analysis tool. This is not in scope of the current contract but will be addressed separately. However, as a first step to realising a new analysis tool, QinetiQ was contracted to identify User and System requirements, and to deliver a Data Analysis Tool Requirements document, which will form a foundation for the development of an analysis tool.

Transition Plan

An important aspect of the overall requirement is the development of a Transition Plan. It is MOD's intention to transition the capability to measure anthropometry to the three services – Army, RN and RAF[3]. Ideally, each of the three services will have the same hardware and software and follow the same protocols in terms of measuring personnel. In this way, it is hoped that an enduring capability will be developed and as a consequence, the measuring of service personnel will become 'business as usual'. In other words, the database will be continually updated and maintained and MOD will never again find itself in the same position as today – having to embark upon another comprehensive tri-service anthropometry survey.

Solution

In response to the requirements identified by MOD, QinetiQ has developed a programme of work, which is currently underway. Key elements of this work programme are described below.

Sampling Strategy

An analysis of the MOD's requirements concluded that a total of 163 body dimensions were required to be measured. These comprised the 92 dimensions measured in the 2006-07 survey plus 53 dimensions required to meet the requirements for new body armour, 12 dimensions required for compatibility with the JACK[4] digital human modelling tool, plus an additional 6 body dimensions

[3] The RAF Centre for Aviation Medicine (RAFCAM) already has a capability and DE&S is indebted to RAFCAM for the support and guidance it has provided to the Tri-service Anthropometry Survey.
[4] JACK is a human modelling tool owned by Siemens with UK distributors SIMSOL Ltd

that were deemed necessary, e.g. Buttock-heel length whilst seated. As 30 dimensions were replicated to provide left and right-hand values, the total requirement resulted in 193 measurements.

Given the requirement to record 193 measurements, QinetiQ calculated that it will be possible to sample 2,875 personnel within the constraints of the available time and funding.

A personnel sampling strategy has been devised to gather sufficient data points in the following primary demographic groups: Sex, Ethnicity and Age; and in the following secondary groups: Service (e.g. Army, RAF, RN), Ranks (e.g. Officers, Others) and Service Groups (e.g. Infantry, Aircrew, Submariners). This strategy was developed by statisticians in the MOD's Defence Statistics organisation and DOSG.

The devised strategy is referred to as a 'Stratified Sample Design'[5]. This requires equal variance in each selected sub-strata[6] for at least one critical parameter. Following guidance in ISO 15535:2012, the coefficient of variance was calculated to select this parameter. Weight was determined to be the critical parameter, as this varied more than any other dimension in the 2006-07 survey. The statistical analysis demonstrated that, if the numbers of participants shown in Table 1 are achieved, all sub-strata have a variance of +/- 4 kg, yielding a 90% level of confidence.

	Total Population	Selected Population	Selected Percent	Sample size	% of sample	% of population
Army	114819	22145	19.3%	1091	37.9%	61.1%
Navy	29937	6899	23.0%	722	25.1%	15.9%
Royal Air Force	35556	10354	29.1%	736	25.6%	18.9%
Royal Marines	7604	1589	20.9%	326	11.3%	4.0%
	187916	40987		2875	100.0%	100.0%
Officer	35885	7221	20.1%	1070	37.2%	19.1%
Other Ranks	152031	33766	22.2%	1805	62.8%	80.9%
	187916	40987		2875	100.0%	100.0%
Male	165991	36265	21.8%	1995	69.4%	88.3%
Female	21925	4722	21.5%	880	30.6%	11.7%
	187916	40987		2875	100.0%	100.0%
17-24	43883	11410	26.0%	898	31.2%	23.4%
25-44	117828	24866	21.1%	1144	39.8%	62.7%
45+	26205	4711	18.0%	833	29.0%	13.9%
	187916	40987		2875	100.0%	100.0%
Infantry	38396	10030	26.1%	711	24.7%	20.4%
Aircrew	4085	1358	33.2%	363	12.6%	2.2%
Submariners	3911	438	11.2%	214	7.4%	2.1%
Other	141524	29161	20.6%	1587	55.2%	75.3%
	187916	40987		2875	100.0%	100.0%
BAME	13334	3023	22.7%	632	22.0%	7.1%
White	168052	36423	21.7%	1875	65.2%	89.4%
Gurkha	4576	991	21.7%	86	3.0%	2.4%
Fijian	1954	550	28.1%	282	9.8%	1.0%
	187916	40987		2875	100.0%	100.0%

Table 1: Stratified Sample Design

During the 2006-07 survey, the overall percentage of females measured was 12.6%, with the relevant services representation being Army (7.7%), RAF (20.9%), and RN (16.5%). Following the stratified sample design will generate substantial percentage increases compared with the previous survey: Overall (30.8%), Army (29.8%), RAF (36.7%) and RN (33.5%).

[5] The 2006-07 anthropometry survey employed a more traditional Simple Random Sample (SRS) design.

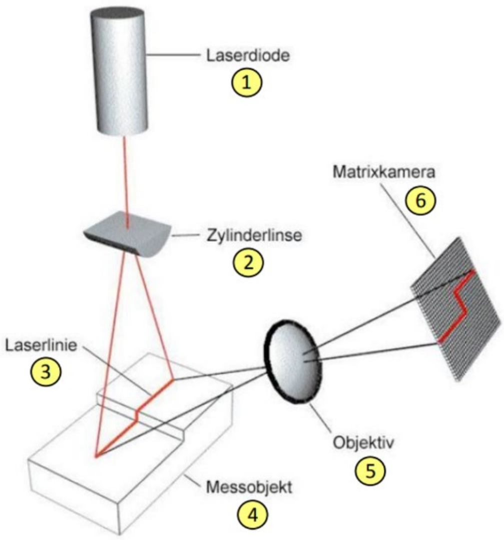
[6] Sub-strata are the primary and secondary groups identified in the previous paragraph.

VITRONICS VITUS Bodyscan 3D Scanners

Two VITRONICS VITUS Bodyscan 3D Scanners have been purchased by MOD. These complement an existing scanner purchased by RAFCAM and will be given to the Army and the RN on the conclusion of this study, thus enabling MOD's transition of the capability to measure anthropometry within the three services.

Figure 2: VITUS Bodyscan 3D Scanner (VITRONIC, 2016)

Figure 2 shows the VITRONICS VITUS Bodyscan 3D Scanner assembled; for scale, the main enclosure is approximately 3m in height. The participant is placed in the centre of the enclosure, the doors are closed to eliminate direct natural light and four scanning sensor heads, one in each corner of the enclosure, travel from the ceiling to the floor taking data measurements. The total scanning process takes approximately 10 seconds.



1	Laser diode	4	Measurement object
2	Cylindrical lens	5	Lens
3	Laser line	6	Matrix camera

Figure 3: VITUS Bodyscan 3D Scanner Laser Light-Section (VITRONIC, 2016)

The data is captured by projecting a horizontal structured light (limited wavelength content, in this case eye-safe near infrared) line onto the body, which is then viewed by a camera system offset from the axis of the illuminating line. The line is seen by the camera as a profile, as seen in Figure 3. Utilising data from a calibration process, the distance of each pixel on the camera from a set datum is converted to an x,y co-ordinate from the central axis of the scanner. The z component is referenced to the floor and acquired from an encoder on the traverse system moving the scanning heads from floor to ceiling. Each scanning head therefore captures a vertical (z) stack of 2D lines described by a series x,y coordinates. The data from the four scanning heads are then combined to provide the complete 3D point cloud describing the surface of the participant, the resultant grid being with a spacing of approximately 1.5mm x 1.5mm between points.

Anthroscan Software

To accompany the purchase of the two scanners, Anthroscan software has been procured for the study. This software allows the capture of pixelated meshes; from very simple ones to highly accurate full figure ones, see Figure 4. A cloud of 3D data points (approx. 1.5 million) is produced and a colour photographic textured map. If MOD wishes to measure some further dimensions, after the study has concluded, these may be captured from the stored meshes.

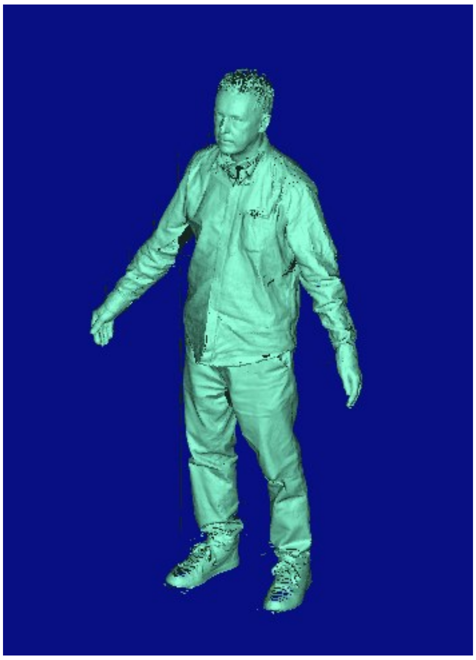


Figure 4: Anthroscan Photographic Textured Map Image

The main use of the Anthroscan software is to create algorithms which extract all the required measurement types from the five 3D scans that will be taken of each participant (2 standing and 3 seated). These will be coded using landmarks and functions within the software to extract specific measurements e.g. breadths, circumferences, contours, heights, lengths, vertical heights, etc., so that all 163 study dimensions are captured.

The Anthroscan software also contains a RAMSIS[7] measurements export wizard for easy export to a Computer Aided Design (CAD) package. Therefore the anthropometry data recorded in this study will support at least two of the main human modelling tools currently available (Jack and RAMSIS).

[7] RAMSIS is a human modelling tool owned by Humanetics Innovative Solutions, Inc.

Military Establishments

After the sampling strategy was derived, an assessment of where the required military personnel were located was performed. The following military establishments were chosen to achieve the numbers required for the study.

British Army	Royal Navy	Royal Air Force
Aldershot	Lympstone	Brize Norton
Andover	Portsmouth	Coningsby
Catterick	Torpoint	Cranwell
Colchester	Yeovil	Halton
Tidworth		Honington
Wattisham		

Table 2: Planned Military Establishment Visits

The number of participants measured is expected to be 75 per scanner per week. Generally, the scanners will be located at the same establishment, though occasionally they will be at two separate bases. A total of 20 weeks' worth of measuring is expected at the 15 establishments. Table 3 illustrates a typical weekly schedule where the scanners are taken to a base and installed.

Monday		Tuesday		Wednesday		Thursday		Friday	
AM	PM	AM	PM	AM	PM	AM	PM	AM	PM
Travel	Setup	Calibration (0.5hrs) Measure	Measure	Measure	Measure	Measure	Measure	Measure	Breakdown Travel

Table 3: Scanner Weekly Schedule

MODREC Protocol

An ethics protocol for the anthropometry survey was developed in conjunction with, and approved by, the Army Scientific Assessment Committee. This protocol was forwarded to the MOD Research Ethics Committee (MODREC). A letter of favourable opinion was received from the MODREC Chief Secretary on January 9th 2023 to enable a Pilot Study to commence.

General Data Protection Regulation (GDPR) and Data Protection Impact Assessment (DPIA)

A DPIA submission provided a comprehensive account of the types of data to be recorded in the study and the uses to which these data will be put. The activity is authorised in law through the explicit consent of participants to have their personal data, which includes information around ethnicity, biometrics and health, collected. Suitable mitigations are required through the use of specialised data storage, data encryption, and pseudonymisation of the data and organisational (contractual clauses, training, vetting) artefacts to achieve security by design and default.

The DPIA was formally approved by the MOD Data Protection Office on January 20th 2023.

Pilot Study

A Pilot Study has been designed to ensure that the scanners and any associated software can capture 3D data from participants and extract the measurements required accurately and consistently and with sufficient fidelity. The study will determine the limits of the Vitus Bodyscan system and Anthroscan software. During the pilot study different materials (colour and reflectivity) will be scanned to determine whether there are any special requirements for underwear (for participants to be scanned in). It has been broken down into two stages:

Firstly, taking traditional measurements of all of the dimensions, to provide a 1-1 correlation between scanned and traditional measures. The techniques stated within ISO 20685-2018 will be used to assess the accuracy of the scanned measurements against the equivalent manual measurements to ensure they are sufficiently close to progress to the second part of the Pilot Study and subsequently, to the Main Survey.

Secondly, a small study of the whole scanning process will be carried out with ten QinetiQ staff, to ensure the end-to-end procedure is understood and any issues are identified early and rectified. The ten participants will therefore be processed by the scan teams (each consisting of four personnel, including at least one female), several times, in a manner (clothing, equipment and procedure) that is identical to the expected Main Survey procedure.

Conclusion

At the time of writing the anthropometry survey is underway, the Pilot Study is due to start in February 2023 with the main survey expected to start in April 2023. It is anticipated that new data to support the development of improved and inclusive body armour will be available in September 2023 and that all data gathering will have completed by the end of 2023. It is planned that the HFI Technical Guide for Anthropometry (Cummings, 2022) will be updated to include the new data in 2024. This will be freely available to all via the MOD's Defence Gateway portal. It is further anticipated that a new anthropometry data analysis tool will be developed and that this will also be freely available to all, although how this will be achieved and when it will be available are yet to be determined.

References

Cummings, R. (2022). Human Factors Integration Technical Guide for Anthropometry: People Size. Version 4.4.

Breeze, J., Lewis, E. A., Fryer, R. (2016). Determining the dimensions of essential medical coverage required by military body armour plates utilising Computed Tomography. International Journal of the Care of the Injured, Vol. 47: 1932-1938.

Breeze, J., Lewis, E. A., Fryer, R., Hepper, A. E., Mahoney, P. F., and Clasper, J. C. (2016). Defining the essential anatomical coverage provided by military body armour against high energy projectiles. Journal of the Royal Army Medical Corps, Vol. 162, Issue 4.

ISO 15535:2012 – General requirements for establishing anthropometric databases (Annex A - Method for estimating the number of subjects needed on a sample).

ISO 20685-1:2018 3-D scanning methodologies for internationally compatible anthropometric databases.

Lewis, E. (2020). Digital mannequins for our diverse Armed Forces. Application for Defence Innovation Fund TLB Ideas Scheme (Round 9) – FY20/21. 27 October 2020.

MOD (2016). Interim Report on the Health Risks to Women in Ground Close Combat Roles WGCC/Interim-Report/10/2016.

Pringle, R. H., Puxley, A. J., Puxley, K. P., Turner, G. M, and Tyrrell, A. K. (2011). Anthropometry Survey of UK Military Personnel 2006-7 (Issue 3). QINETIQ/07/01821/3.0.

Protecting those who protect us: Women in the Armed Forces from Recruitment to Civilian Life - Defence Committee - House of Commons (parliament.uk).

VITRONIC (2016). VITUS Bodyscan Operating Manual Version 1.4 - 22.01.2016.

Developing an Explainable AI Recommender System

Prabjot Kandola and Chris Baber

School of Computer Science, University of Birmingham

ABSTRACT

We used a theoretical framework of human-centred explainable artificial intelligence (XAI) as the basis for design of a recommender system. We evaluated the recommender through a user trial. Our primary measures were the degree to which users agreed with the recommendations and the degree to which user decisions changed following the interaction. We demonstrate that, interacting with the recommender system, resulted in users having a clearer understanding of the features that contribute to their decision (even if they did not always agree with the recommender system's decision or change the decision). We argue that the design illustrates the XAI framework and supports the proposal that explanation involves a two-stage dialogue.

KEYWORDS

Explainable AI, Recommender Systems, Travel planners

Introduction

Explainable AI (XAI) is a set of processes and methods to allow humans to comprehend the output of AI systems. Often these approaches emphasise the 'interpretability' of the model, i.e., how easily humans can understand the underlying model used by the AI system. An alternative approach is concerned with 'explainability' where the AI system is explaining its results, often in terms of the features which may have led to a particular output (Kaur et al., 2022; Erasmus et al., 2020). However, both approaches have a tendency to be AI-centric rather than human-centred, i.e., the approaches assume that the human needs to *understand* what the AI system has done and why it has done this. Such understanding need not be important to many forms of explanation (Mueller et al., 2019). Adadi and Berrada (2019) proposed four reasons as to why people need explanations from AI systems.

- Explain to justify: the AI system must justify why that explanation resulted;
- Explain to control: the AI system provides sufficient information for the user to identify and correct errors;
- Explain to improve: the user is able to correct the model that the AI system is using, so that the performance of the AI system can be improved;
- Explain to discover: the user is able discover the beliefs that the AI system is using, perhaps through testing with counter-factual examples.

When presented with XAI tools, there can be a tendency for users to over-trust such tools (Kaur et al., 2020) or the visualizations that are used (Hohman et al., 2020). Mueller et al. (2019) concluded that an explanation needs to focus on global rather than local explanations, on the performance of the user and encourage the user to reflect on their own interpretation of the output of the AI system.

In other words, the purpose of 'explanation' should not simply be to train the user to understand what the AI system is doing but to enable the user to better integrate the output of the AI system into their decision-making. This presents a departure from AI-centric approaches but faces two fundamental barriers:

(1) There are no universal criteria as to what defines an adequate explanation from an AI system. Therefore, AI system developers have no standard definition to follow when developing explanations;

(2) Even if there were universal criteria, these might not be applicable to users of the AI system for all contexts of use.

In previous work, we argued that an explanation ought not to be solely the concern of the direct user of the AI system but with anyone affected by the AI system, i.e., those who program the system as well as analysts who interpret its output and other stakeholders affected by the decisions based on the AI system's output. This is a tall order, but one that a human-centred approach could help address. To do this, we have proposed a framework that specifies the kind (s) of knowledge an AI system should provide so the 'Explanation' would be both 'interpretable' and 'explainable' to all stakeholders (either through their direct interaction with the AI system or through indirect actions, i.e., where the output of the AI system is communicated by an intermediary). More simply, XAI systems only focus on decision relevant features and the definition of 'relevance' that has been applied.

A model of Explanation

'Explanation' involves common ground in which two parties are able to align features to which they attend *and* the relevance that they apply to these features. We use this proposal as the basis for designing a recommender system. Baber et al. (2020, 2021) suggest that much of the prior work in XAI systems 'provides an output only at the level of features. From this the user has to infer Relevance by making assumptions as to the beliefs that could have led to that output. But, as the reasoning applied by the human is likely to differ from that of the AI system, such inference is not guaranteed to be an accurate reflection of how the AI system reached its decision.' That is, users are likely to have a different understanding (which may be due to demographics such as age, gender, education or salary) to the system in terms of 'why' a feature may have been chosen for a particular decision, or 'what' features could be used to reach a particular decision. Therefore, if the user disagrees with a decision, the user will have to infer what other features they would need to choose from to receive a recommendation which is more closely related to what they would want. In this case, the goal of explanation would be to align the features to which they attend and the type of relevance that they apply to these features. These assumptions are presented in figure 1. This framework suggests that an explanation involves an 'agreement on features (in data sets or a situation) to which the explainer and explainee attend and an agreement on why these features are relevant (this proposes three levels i.e., 'cluster' in which a group of features will typically occur together, a 'belief' which is a reason as to why these clusters occur, and policy which justifies the belief related to this action.

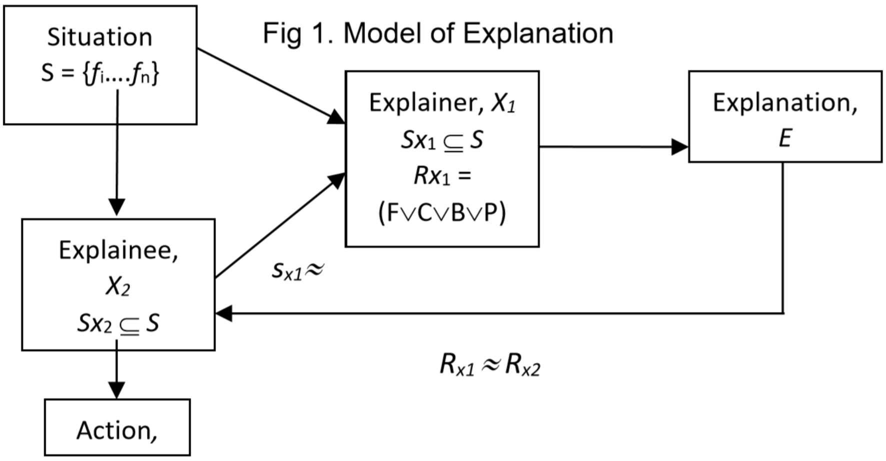


Figure 1: Framework for Human-centred XAI (Baber et al., 2021). A situation, S, has a set of features, {fi….fn}, which can be described symbolically, using words, numbers, pictures, etc. For example a situation might be the user choosing features (i.e. price, time) they believe are important when travelling from University of Birmingham to UoB. The Explainer is the XAI system's set of features which contribute to an Explanation. The Explainee is the users and includes the set of features to which they attend. Action is the action which could be taken by the user in light of the explanation.

An Explainable Journey Recommender System

In this paper, we develop a recommender system based on the model presented in figure 1There were two challenges in developing this recommendation system in order to meet the criteria suggested in the framework:

(1) The set of features which the Explainer X1 attends should overlap the set of features used by the Explainee X2

(2) Ensure the Explainer and Explainee can agree on what features define a situation i.e. define 'relevance'

The recommender system suggests routes for a user to take when travelling from University of Birmingham to the City Centre or vice versa. For the user to receive a recommendation they need to indicate features they believe are important when make a travel decision. Features could include 'price' or 'time', e.g., if the user chose 'time' and ranked this as '1', the recommendation would suggest taking an 'Uber' since it faster than Cycling or Walking.

The first challenge of the recommender system is for the Explainer (recommendation system) and Explainee (user) to attend to the same features. The second challenge of this system is for the Explainee and Explainer to have a similar concept of relevance which can achieved through dialogue between Explainer and Explainee (conducted, in this instance, using a chatbot).

Defining Features

The interaction commences with the user selecting a destination for the journey (figure 2). We use a user interface design familiar from ticket vending machines. The defines the scope of the Situation for the recommender system.

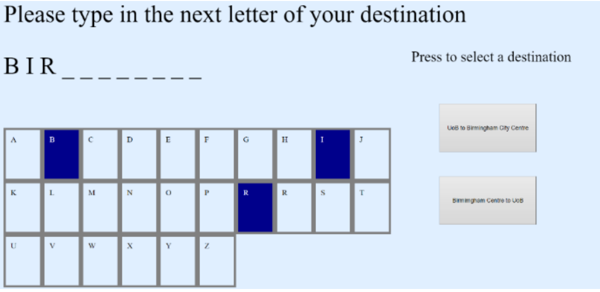


Figure 2: Defining a destination

As indicated in figure 1, a situation, S, has a set of features, {fi….fn}. In our model, we assume that the Situation also includes the constraints that define an acceptable decision. For this, we invite users to select ‘features’ that they believe to be relevant to their choice of journey type.

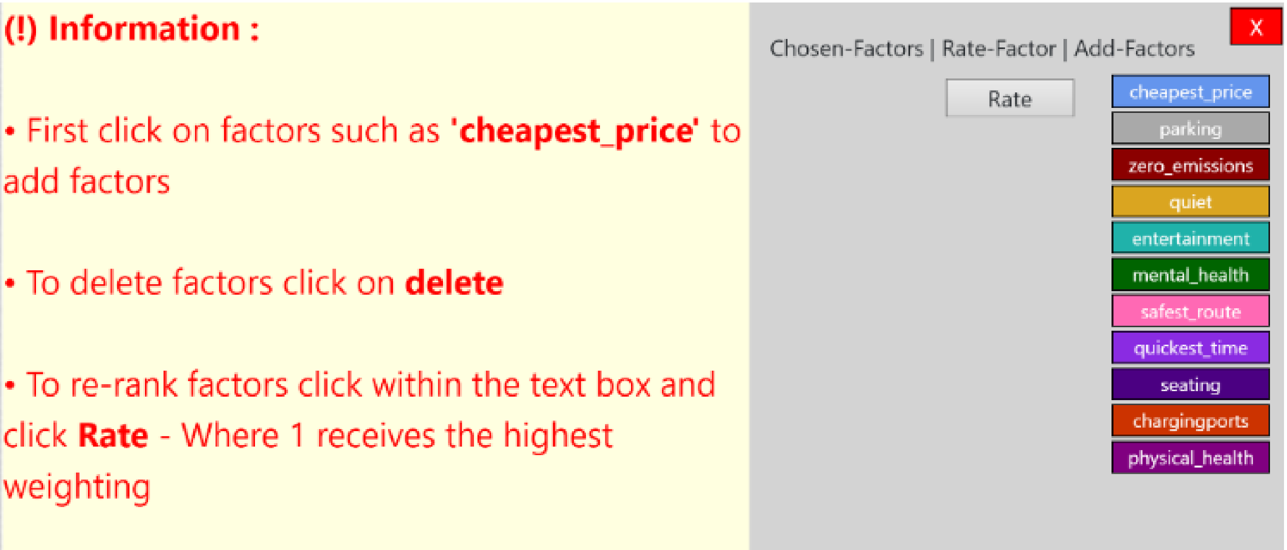


Figure 3: Defining Features

As figure 3 shows, users can select from a set of features (derived from an initial study with transport users). This initial set included {timing, price, emissions, congestion, capacity, number of changes, health, entertainment, charging ports, seating, safety, quiet, parking} can be expanded during user trials where participants offer additional features. The weighting of each selected feature is defined as a ratio of the number features selected such that the magnitude decreases, i.e., if the user selects 3 features then this produces weights of 0.5, 0.35, 0.15 (figure 4)

Figure 4: Calculating the weight of each feature

A pre-defined SQL database scores all features for each mode of transport {bus, taxi, car, train, walking, cycling). From this, the user weighting is combined with the mode of transport scoring. For example, the selected features map to the mode ‘car’ as shown in figure 5.

user_transport_weights1

cheapest_price	zero_emissions	entertainment
0.15	0.35	0.50

*

user_transport_weights1 (This is the car route)

cheapest_price	zero_emissions	entertainment
0.60	0.25	1.00

Figure 5: Ranking the features. In this case, the overall rating of mode: car is defined as: Price (0.15* 0.6) = 0.09 + Emissions (0.35*0.25) = 0.0875 + Entertainment (0.5*1) = 0.5 = 0.68.

Applying these features to the other modes of transport allows us to create a ranked list based on these ratings (figure 6).

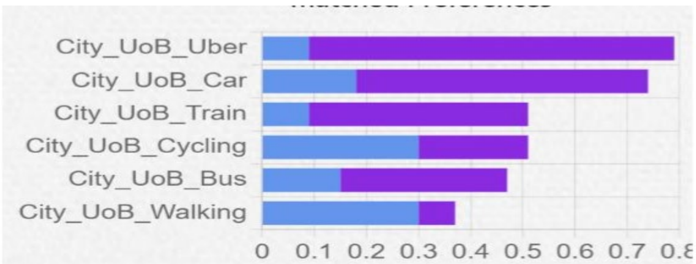


Figure 6: Rank ordering of all journey options

In addition to ranking transport relative to user-defined features, we also implement an algorithm which uses the features chosen by the user, as well as the features 'zero emissions' and 'physical health' to provide a recommendation. The 'zero emissions' and 'physical health' would always be given a higher ranking than the features selected by the user. While this process does *not* use Artificial Intelligence, we felt that it was sufficienty opaque for users to have difficulty in interpreting the recommendation and how it was derived. In other words, the purpose of this activity was not to simulate AI per se but to produce a recommendation that required explanation.

The recommender system implements a chatbot which makes the user aware of their features and how they relate to their routes as well as provides justifications into why specific decisions have been made to then nudge the user into changing their idea of what a good decision is based on the ideal recommendations (figure 7). This is where the system and the user work together to identify what features they believe are important and what a good decision is.

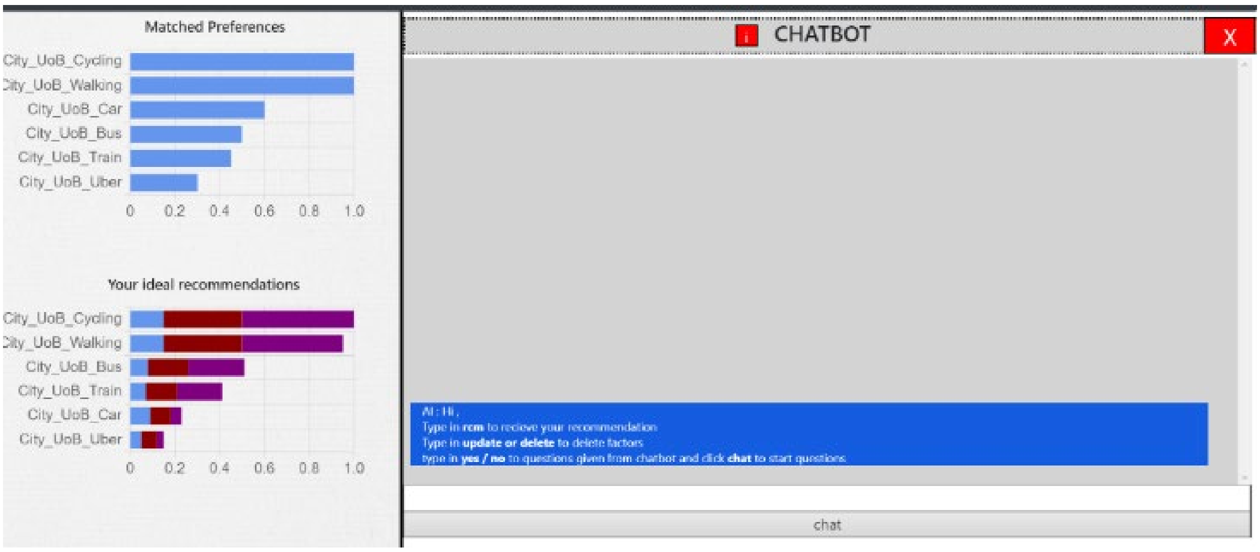


Figure 7: User interface showing preferences, 'ideal recommendations' and chatbot

Figure 7 shows the user interface with which the user can interact. There are three actions that the user can take: the user will agree with the explanation and choose a particular route; the user will disagree (not same) with the explanation, and ‘Person 1’ would be given new recommended features and then go to the action; the user can accept the recommendation and is shown a map with detailed instructions of the journey (figure 8).

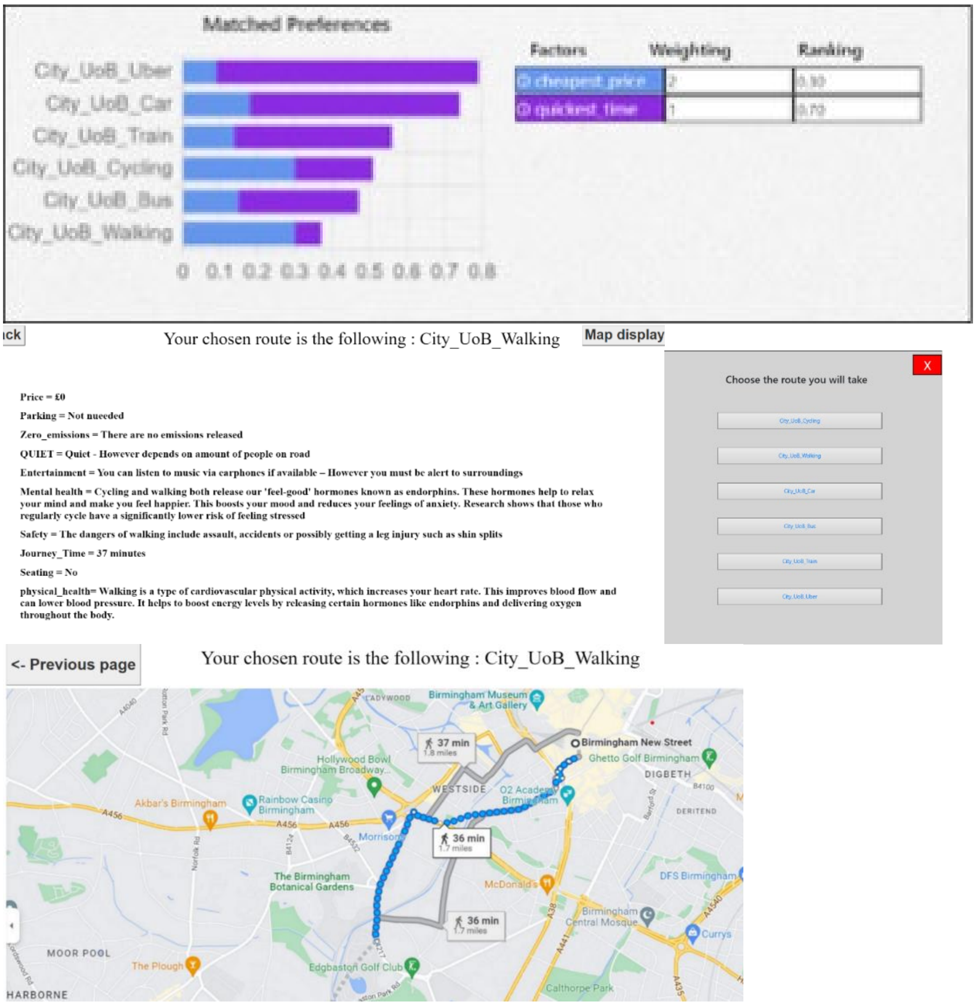


Figure 8: Journey plan output by the recommender system

Evaluation

20 participants were involved within this study, participants were either current or graduates from several universities. The age of participants ranged from 21 – 30. We accept that this produces a homogenous sample but propose that this makes it easier to aggregate the results. Future work could explore different user groups through more stratified sampling.

Participants were asked to interact with the recommender system in order to define a journey. They were asked, before the interaction began, what factors they normally consider when planning a journey. We used this to define the baseline against which we could compare the set of features that were considered following the interaction. As they interacted with the recommender system, we asked as a form of Cognitive Walkthrough them to articulate their impressions of the system’s operation, whether they understood its recommendations, and whether the interaction had altered their choice of features or decision on journey type.

Prior to the interaction, the main features participants considered were ‘price’ and ‘time’. This agrees with prior studies which states ‘people typically only mention one or two features’ [4]. For participants, ‘time’ was associated with their experience of going into university such as arriving to a lecture on time or attending a meeting. Some users also noted ‘time’ as important since they don’t like to ‘waste time’ during the day, this could be because they have other activities such as the ‘gym’, ‘university work’ or wanting to go out with ‘friends and family’. Some users noted ‘weather’ as something they would consider; it was found that this would affect their travelling arrangements or the time they might leave their house. The type of transport they would take was also brought up, for example some users would often talk about ‘train’ or ‘uber.

Following the interaction, participants reported more features as relevant to their decision. Figure 9 shows the effect of interacting with the interactive chart or the chatbot on the number of features mentioned by participants.

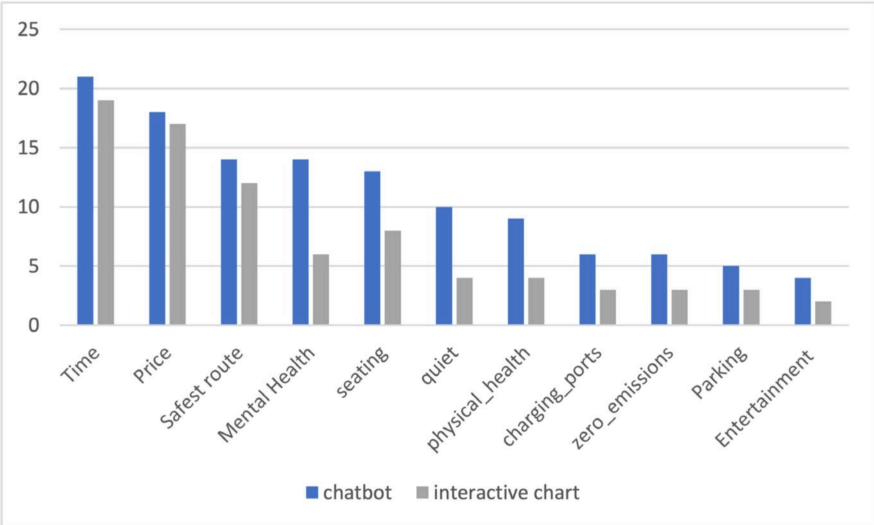


Figure 9: Count of features mentioned by participants

Combining participant response, we constructed a concept map(figure 10).

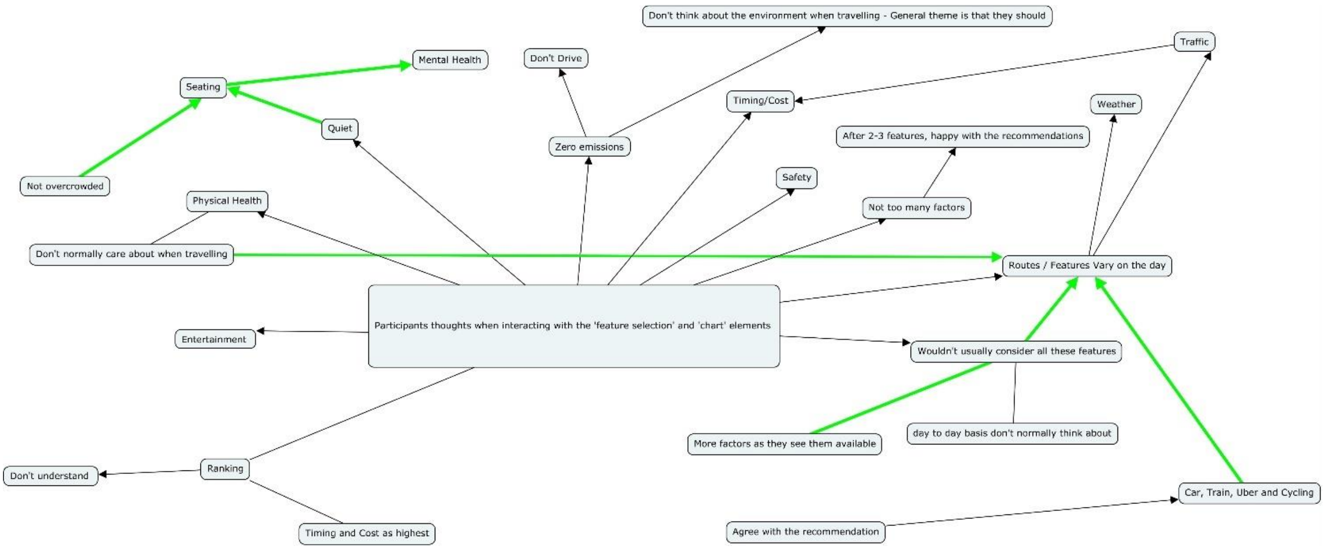


Figure 10: Concept map from participants following interaction with the recommender system

On whole, participants preferred their own ‘matched preferences’. However, the majority (17/20) of participants believed that the ‘ideal recommendations’ derived from physical health ‘made sense’ and these should be routes they should take. The reasons given for not including these features or taking the ideal routes was because they were ‘too lazy’, ‘cycling would take too long’ or ‘walking would take too long'. Participants were less inclined to include zero emissions as a feature in their decisions (mainly as participants did not drive to university).

Furthermore, the chatbot helped participants understand 'why' features were chosen for their particular route, and once they understood the reasoning for this feature, they could then re-rank and alter this within their charts to receive a recommendation more closely related to what they would want, e.g., some participants did not understand how safety would lead to a higher rating for walking and thought a car to be safer. However, after understanding the XAI' s reasoning they agreed with the reasoning for this. When the chatbot gave its justifications for zero emissions it was found it was not enough to persuade the user into changing their minds, participants were hesitant to include zero emissions as they understood this would result in a change in the order of their recommendations where 'cycling' or 'walking' would be placed higher. This is preferable as participants now understood why and how features led to specific recommendations rather than having to 'inferred' this themselves.

To conclude, the recommendation system did not force users to change their minds or alter their choice. In this case, it was not particularly useful to 'nudge' their choices. However, this was not the primary intention of the project. Rather, we have demonstrated how a design for a recommender system can be developed from our XAI framework and that interacting with this recommender system helped users to elaborate on the features that inform their choice, and to understand how the recommender system has produced its recommendation – both of which we believe are integral to developing XAI.

References

Adadi, A. and Berrada, M., 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Access, 6, 52138-52160

Baber, C., McCormick, E., & Apperley, I., 2021, A human-centered process model for explainable AI. In: Naturalistic decision making and resilience engineering symposium 2021, France.

Baber, C. McCormick, E. and Apperley. I., 2020, A framework for explainable AI, *Proceedings of Institute of Ergonomics and Human Factors*

Erasmus, A., Brunet, T.D. and Fisher, E., 2021. What is interpretability? *Philosophy & Technology*, *34*, 833-862.

Hohman, F., Wongsuphasawat, K., Kery, M. and Patel, K., 2020. Understanding and Visualizing Data Iteration in Machine Learning. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, New York: ACM, 1-13.

Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H. and Wortman Vaughan, J., 2020, April. Interpreting interpretability: understanding data scientists' use of interpretability tools for machine learning, *Proceedings of the 2020 CHI conference on human factors in computing systems*, New York: ACM, 1-14.

Mueller, S.T., Veinott, E.S., Hoffman, R.R., Klein, G., Alam, L., Mamun, T. and Clancey, W.J., 2021. Principles of explanation in human-AI systems. *arXiv preprint arXiv:2102.04972*.

Mueller, S.T., Hoffman, R.R., Clancey, W., Emrey, A. and Klein, G., 2019. Explanation in human-AI systems: A literature meta-review, synopsis of key ideas and publications, and bibliography for explainable AI. *arXiv preprint arXiv:1902.01876*.

How the Accuracy of Interactive Voice Assistants Affect Perceived Trust

Wenhu Zhang and Chris Baber

School of Computer Science, University of Birmingham

ABSTRACT

We asked 30 participants to ask questions of an Interactive Voice Assistant (IVA) which we had modified to provide different levels of accuracy in its answers. The levels of accuracy were low (55%) or high (80%). We also told users what level of accuracy to expect (60% or 100%). This produced a set of 6 combinations of actual accuracy with expected accuracy (including the condition when we did not tell the users which level of accuracy to expect). As expected, when users experience a more reliable IVA (i.e., 80% vs. 55%) their rating of trust is higher, and when actual an IVA with high accuracy and they are expecting accuracy to be high, then their trust rating is higher still. However, *expected* accuracy seems to outweigh actual accuracy, particularly when the actual performance is less than expected. Counter intuitively, this suggests that participants were not able to judge the actual accuracy of the IVA but relied on the *expected* accuracy.

KEYWORDS

Trust, Interactive Voice Assistants

Introduction

The aim of this study was to explore how Trust if affected by accuracy when people speak to an Interactive Voice Assistant. Specifically, we were interested in how the level of accuracy that users expect from the device compares with the level of accuracy that they experience when using the device. The point of manipulating expected accuracy (in addition to the actual accuracy) was to explore interaction effects on trust arising from expected and actual accuracy. For example, if users expect the device to have low accuracy and it performs very well, does this have a positive impact on their trust in the device.

Research into human trust in automation (TiA)began in earnest with pioneering work of Muir (1994) and Lee and Moray (1992). When it comes to perceiving the level of trust in an automation, users have few criteria to judge it other than its stated performance, their observations during actual use and their own knowledge domain. This raises the question of whether users can determine the reliability or the accuracy of the automation with which they are interacting, or whether they rely on prior experience of their interactions, or simply revert to trusting performance claims provided to them. For example, would people trust an automation more if they were told that it was 90% accurate on retained data rather than 60%? If more trusting, does its stated accuracy when they actually use the automation still affect that trust?

There remains a lack of a universally accepted model of TiA, perhaps because trust varies with context and type of automation. To some extent, this assumes that ‘trust’ is dispositional, i.e., trust is a subjective response to the performance of automation. In this paper, we adopt Mayer and Davis’s (1995) model of trust (figure 1). In this model, trust arises from a combination of the user’s propensity to trust and factors which affect perceived trustworthiness of the agent (human or automation) they interact with. The experience of using automation could lead to changes in user response (either physiologically or behaviourally), particularly when the agent behaves in unexpected ways. This implies that, while the model appears to be dispositional, it has a strong activity component.

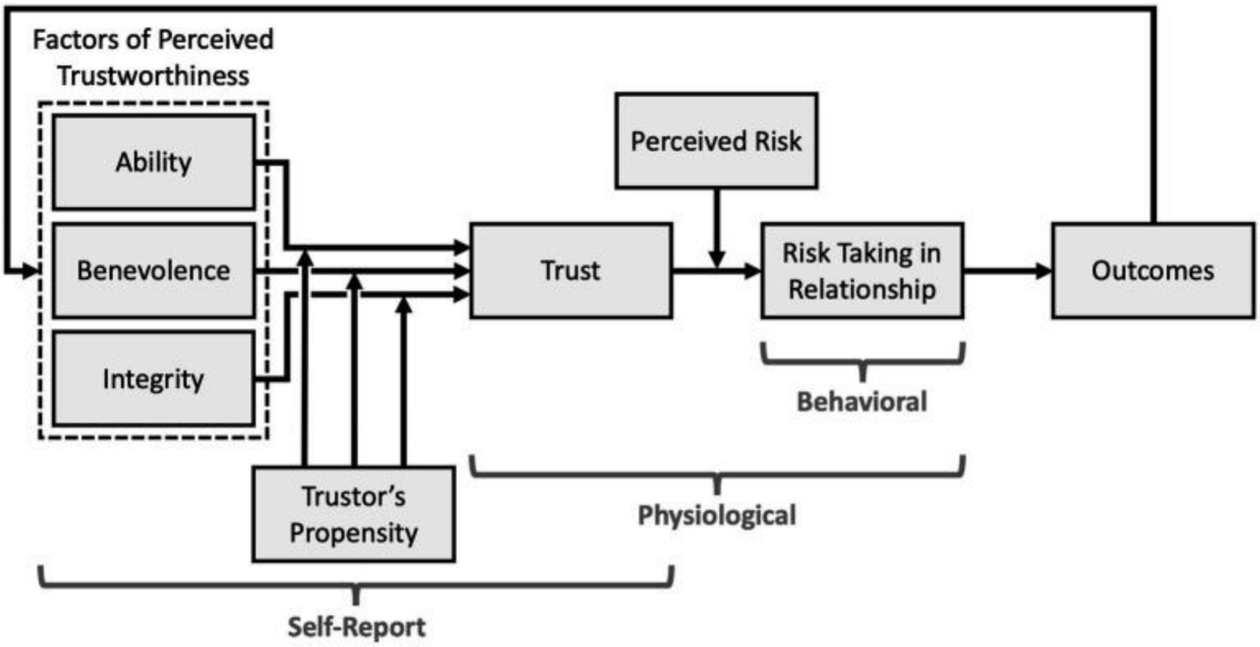


Figure 1: Mayer and Davis' (1995) model of trust

Kohn et al. (2021) consider Meyer and Davis' (1995) model in terms of the ways in which trust could be measured (table 1).

Table 1: Relating types of trust to measures

Trust Type	Trust Process Step	Measure	Experiment step
Factors of perceived trustworthiness	Perception of the system's trustworthiness-related characteristics	Self-report from user	Before / during interaction
Trustor's propensity	Effects of individual's traits	Self-report from user	Before interaction
Trust	Trust stance or attitude that exists during interactions and influenced by feedback	Self-report; Physiological measures	During / after interaction
Perceived risk	Effects of individual's understanding of situation	N/A	Pre-interaction in environmental situation
Risk-taking	Trust behaviour expressed during interactions	Behaviour	During / after interaction
Outcomes	System accuracy and user trust	N/A	After interaction

Method

Participants

We recruited 30 participants for this study (mean age 26 (± 15); 9 female). Participants had no previous experience of using Interactive Voice Assistants.

Equipment

We used Google's Dialogflow to create a bespoke Interactive Voice Assistant (IVA) and participants interacted with a Bose Soundlink Revolve II (figure 2) to present spoken response to questions.

Figure 2: Bose Soundlink revolve II

We modified the performance of the IVA so that it provided correct answers to either 55% or 80% of the questions. The correct was specified in our question set; an incorrect answer was a random choice of answer from the question set.

Measures

We used Jian et al's. (2000) Human-Automation Trust Checklist to gather subjective responses from participants. This is a scale that has been widely used in the literature. This checklist includes 12 questions which reflect a range of attitudes which can affect the perception of trust in automation. The checklist has separate sets of questions for Distrust and Trust (5 for distrust and 7 for trust). While it is popular means of evaluating trust, Gutzwiller et al. (2019) urge caution in its use because it could be skewed towards positive ratings. When the survey is completed in the original order of statements and with the original rating scales used, participants tended to produce higher ratings of trust than when the survey was presented in other configurations (albeit the effect was quite small). As a simple expedient to minimise the potential bias, we subtracted the median ratings for the Trust questions from those of the Distrust questions (on the assumption that this was indicate propensity to trust the IVA). However, this means that the application of the Human-Automation Trust Checklist is different from the original intentions of its developers.

Procedure

We assumed that accuracy would primarily be affected by the IVA misrecognising a complete sentence, rather than individual words. In this manner, we could compare the impact of medium and high accuracy of user trust. Additionally, we told participants that the IVA had an accuracy of either 60% or 100%.

The combination of experienced and expected accuracy produced six experimental conditions, as shown in table 1. The order in which participants experienced the conditions was counter-balanced using Latin squares. We repeated the conditions for each question set. Examples of questions are shown in table 1.

Table 1: Examples of questions used for the experiment

Question Set1	**Answer**
What is 20% of 80?	Twenty percent of eighty is sixteen
What does "Sociable"mean?	Sociable means to have a harmonious relationship with everyone and get along well
How many days are there in a year?	There are three hundred and sixty-five days in a year
How many kilometres are in a mile?	A mile is about 1.6 kilometres
How is the road to work?	Good road to work
How much is five plus seven?	Five plus seven equals twelve
What day is it today?	Today Thursday
How is the traffic situation?	Good traffic conditions nearby
How's the weather today?	It's sunny now, the temperature is 30 degrees
Question Set2	**Answer**
Add meeting to my calendar	All meetings added to calendar
Adjust the temperature of the bedroom air conditioner to 26 degrees	The bedroom air conditioner has been adjusted to 26 degrees
Add milk to my shopping list	Added milk to shopping list
What is the current volume?	The current volume is 50%
When is my first meeting today?	First meeting today at 3pm
What questions can I ask you?	You can ask me for help with information and daily tasks.
What is the battery level of my speaker?	Battery is 66%
Remind me to call mom every Sunday	OK, the reminder is set to start this Sunday
When will sunrise tomorrow?	Sunrise tomorrow is 5:30
How is the S&P 500 performing?	S&P 500 shares rose to 3998.95 today, up 0.99%

Results

Initial analysis of the responses to the checklist was performed, for each question set and across each experimental condition, using Cronbach's alpha and Kaiser-Meyer-Olkin sampling adequacy. The results, shown in table 2, indicate high levels of agreement within the checklists across all conditions. From this, we assumed that it would be appropriate to merge responses to question sets 1 and 2 for subsequent analysis.

Table 2: Agreement of participants in their trust ratings

Accuracy (expected)	60%	100%	60%	100%	Not told	Not told
Accuracy(actual)	55%	55%	80%	80%	55%	80%
Condition	A	B	C	D	E	F
Cronbach's alpha	0.847	0.8295	0.873	0.8325	0.855	0.847
KMO	0.7675	0.7965	0.8305	0.702	0.7705	0.7675

We analysed the median rating of the seven 'Trust' questions (figure 1). A Friedman Analysis of Variance, calculated using R, showed a significant main effect of Condition [x2 (5) = 109.5, p<0.0001]. With the exception of C x F, all post-hoc comparisons (using Wilcoxon Signed Ranks test) were significant at the 5% level. The highest rating of trust was for condition D (in which expected and actual accuracy were high).

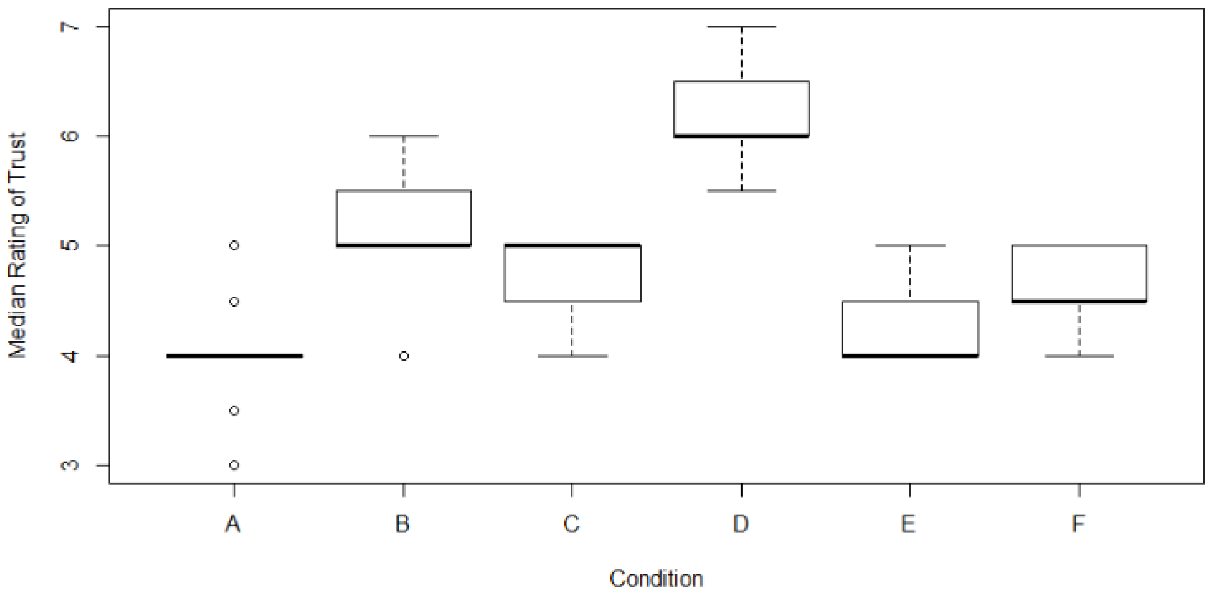


Figure 1: Rating of 'Trust' between the different conditions

Conclusions

We compared trust ratings when using an IVA under different manipulations. We were interested in how expected accuracy or actual accuracy affected these ratings. The results in figure 1 can be grouped into four observations.

First, conditions A (low expected + low actual accuracy) and E (no expected + low actual accuracy) are similar. This suggests that participant could detect when the actual accuracy of the IVA was low.

Second, condition D (high expected + high actual accuracy) is significantly different to the other conditions (at p<0.05 using Wilcoxon pairwise, post-hoc tests). This suggests that participants were positively influenced by high expected and high actual accuracy. While this is to be expected, it suggests that (coupled with observation 1) that participants were moderating their trust ratings in a predictable manner.

Third, when participants have not been told the accuracy of the IVA (conditions E and F), there is no significant difference in trust in terms of actual accuracy. This was surprising, given observation 1, because it suggests that detection of actual accuracy is not as simple as we might assume. It might be that we had too few questions in our question sets so that participants did not have long enough exposure to the IVA to form an opinion of its accuracy. An alternative explanation that, without being given expected accuracy, participants begin with a low expected accuracy (possibly lower than the one we provided) because they are not sure whether the IVA would recognise their speech. This was mentioned by a few of the participants. In this case, rather than the 'trust' being in the IVA it would be based on whether the IVA would respond to the participants (and, the implication here is that the participants might place the locus of any performance failures on themselves and their inability to get the IVA to work as much as on the failure of the IVA to respond to them).

Fourth, the trust rating conditions B (high expected + low actual accuracy), C (low expected + low actual accuracy), and F (no expected + high actual accuracy) show no difference. This also suggests that rating of trust is moderated by actual accuracy (B and C) but that there is also an a priori assumption that the IVA will have low performance.

Our findings bring a number of practical implications for research on Human-Computer interaction and trust in automation, thereby increasing trust in automation. Firstly, our findings show that designers of automation must express their expectations of automation accurately and responsibly, because only in this way can users determine the extent to which they can trust automation before and after interaction. Furthermore, we have found that users focus on the process of interacting with automation even when their interaction with it is limited to a few tasks. For example, if the actual use is very different from the automation training, i.e., if the stated accuracy does not reflect well the accuracy observed in its actual use. However, this small amount of task feedback is not indicative of the average performance of the automation in actual use. Therefore, it is important for automation designers to communicate well the uncertainty that the automation will complete correctly based on a small number of tasks. In this way, even if the user observes low performance on the successful completion of the first few tasks when using the automation, this will not lead to a false distrust of an automation. All of the above are implications of the project's findings for automated trust research.

Finally, our work highlights that in order to understand the interactions between humans and the different components of automated work, more experimentation is needed to enable work on the interpretable aspects of machine learning within automation to go beyond the current focus on its models themselves.

References

Gutzwiller, R.S., Chiou, E.K., Craig, S.D., Lewis, C.M., Lematta, G.J. and Hsiung, C.P., 2019, November. Positive bias in the 'Trust in Automated Systems Survey'? An examination of the Jian et al.(2000) scale. In *Proceedings of the Human Factors and Ergonomics Society annual meeting*(Vol. 63, No. 1, pp. 217-221). Sage CA: Los Angeles, CA: SAGE Publications.

Jian, J., Bisantz, A., & Drury, C. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, *4*(1), 53–71.

Kohn, S.C., de Visser, E.J., Wiese, E., Lee, Y.C. and Shaw, T.H., 2021. Measurement of trust in automation: A narrative review and reference guide. *Frontiers in psychology*, *12*, p.604977.

Lee, J. and Moray, N., 1992. Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, *35*(10), pp.1243-1270.

Mayer, R.C., Davis, J.H. and Schoorman, F.D., 1995. An integrative model of organizational trust. *Academy of management review*, *20*(3), pp.709-734.

Muir, B.M., 1994. Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, *37*(11), pp.1905-1922.

Appendices A Jiang et al. Human-Automation Trust Checklist

Checklist for Trust between People and Automation

Below is a list of statement for evaluating trust between people and automation. There are several scales for you to rate intensity of your feeling of trust, or your impression of the system while operating a machine. Please mark an "x" on each line at the point which best describes your feeling or your impression.

(Note: not at all=1; extremely=7)

1 The system is deceptive

1 2 3 4 5 6 7

2 The system behaves in an underhanded manner

1 2 3 4 5 6 7

3 I am suspicious of the system's intent, action, or outputs

1 2 3 4 5 6 7

4 I am wary of the system

1 2 3 4 5 6 7

5 The system's actions will have a harmful or injurious outcome

1 2 3 4 5 6 7

6 I am confident in the system

1 2 3 4 5 6 7

7 The system provides security

1 2 3 4 5 6 7

8 The system has integrity

1 2 3 4 5 6 7

9 The system is dependable

1 2 3 4 5 6 7

10 The system is reliable

1 2 3 4 5 6 7

11 I can trust the system

1 2 3 4 5 6 7

12 I am familiar with the system

1 2 3 4 5 6 7

Psychophysiological coherence training reduces pilots' perceived stress in flight operations

Laurie Marsman[1,2], Jingyi Zhang[1] & Wen-Chin Li[1]

[1] Safety and Human Factors in Aviation, SATM, Cranfield University, UK, [2] Netherlands Aerospace Centre (NLR)

SUMMARY

The present paper reports the results of a four-week study assessing the relationship between psychophysiological coherence training, which aims to improve the synchronisation between one's physiological rhythms leading to a positive emotional state, and perceived stress levels in commercial air transport pilots. Next to three self-report stress questionnaire measurements, qualitative data were gathered as well to gain more insights into the training effects. Results show significant reductions in the flight crew's perceived stress levels during the four-week psychophysiological coherence practice period. Finally, the results are discussed.

KEYWORDS

Heart Rate Variability, Resilience, Aviation, Safety, Human Factors

Introduction

Fatigue and stress are well-documented issues within the field of aviation, with far-reaching consequences for all personnel, including pilots. Effective strategies are needed to mitigate the risk these adverse physiological and psychological states pose for aviation safety. One potential solution is fostering pilots' psychophysiological coherence, which is defined as a positive emotional state in which physiological rhythms between the heart and brain become synchronised (Field, 2018), leading to a sense of mental clarity. Such a state is characterised by a low-frequency heart-rate variability (HRV) of around 0.1 Hz (McCraty, 2022). One technique to achieve higher coherence levels is the Quick Coherence Technique (QCT), which instructs participants to breathe at a calm, comfortable pace while accessing positive feelings and focusing attention on the heart (Henriques et al., 2011).

It is possible to visualise one's HRV through biofeedback methods, which depict bodily functions to gain more awareness of one's body. Eventually, the goal is to achieve a state of coherence without biofeedback, ensuing better cognitive performance, and fewer negative emotional states like stress. Currently, psychophysiological coherence training is not used for flight crew. However, there are ample indications that this group can significantly benefit from such training, as flight crews tend to suffer from adverse mental states which can impact their performance and thus flight safety. Hence, this paper investigates the impact of coherence training on flight crew's perceived stress.

Methods

Participants include staff of an international airline (N=28) based at a London airport, all of which signed up voluntarily. No inclusion or exclusion criteria were used. Nine participants (32.14%) are captains, seventeen (60.17%) are first-officers and two (7.14%) are Human Factors experts within the airline. Ethical approval was provided by the Cranfield University Research Ethics System (CURES/16348/2022) and all participants provided their informed consent.

Each participant took part in a one-day training course in which principles of psychophysiological coherence were explained and participants were introduced to the non-intrusive HeartMath Inner Balance biofeedback device used during the study. During the training day, stress levels were assessed with the Perceived Stress Scale (PSS) as a pre-measurement. The PSS is a highly validated scale in a variety of samples (Lee, 2012), and contains ten items which are answered on a scale of 0 (Never) to 4 (Very often). Thus, the total PSS score lies between 0-40, with higher scores indicating higher stress levels. After the training day, participants completed four weeks of independent coherence practice using QCT. Using a within-subjects design, stress levels were assessed in the middle of training and at the end of training, resulting in three PSS measurements. Participants were encouraged to practice the QCT as often as possible with a recommended six sessions a day, three of which are baseline measurements, the other three QCT practice sessions.

Results

Firstly, it was verified that QCT practice (M = 3.715) resulted in significantly higher coherence scores than baseline measurements (M = 1.506) over the four-week period, using a repeated-measures ANOVA ($F(1, 21) = 102.792$, $p < .001$, $\eta^2 = .830$). Seven participants (25%) did not complete all PSS-measurements; hence they are removed from analysis. For the remaining 21 participants, a repeated measures ANOVA was conducted with the PSS scores as dependent variable and time (pre, during, post) as within-subject factor, showing a strong significant main effect of time ($F(2, 18) = 33.996$, $p < .001$, $\eta^2 = .791$). Univariate follow-up shows strong significant main effects of time on the during versus pre-measurement ($F(1, 19) = 26.638$, $p < .001$, $\eta^2 = .584$) and on the post versus pre-measurement ($F(1, 19) = 65.071$, $p < .001$, $\eta^2 = .774$). The pre, during, and post-measurement EM means are 19.850, 13.550, and 13.400 respectively.

Discussion

The present results show significant reductions in flight crew's perceived stress levels during the four-week period of coherence practice, indicating that coherence training is indeed effective in lowering perceived stress levels in commercial air transport pilots, which may positively affect their performance on the flight deck. It should be noted, however, that neither participants' coherence progress, nor the number of completed practice sessions are considered in this paper. Moreover, participants registered for the study voluntarily, highlighting a potential self-selection bias which could have influenced the results.

Regardless, participants provided detailed qualitative feedback after practice sessions, in which they report "I'm feeling slightly calmer after QCT sessions" and "I am beginning to see a marked improvement to my stress levels and situations I would normally find stressful, I'm able to manage more thoughtfully", suggesting that coherence practice may make a marked difference in participants' experienced stress levels.

Importantly, more research is needed to establish a more definite link between coherence training and flight crew's stress levels and subsequent performance, preferably including control groups, and more detailed stress, coherence progress, and in-flight performance assessments. The social importance of such research cannot be understated: The impact of pilots' adverse mental states on pilot performance and flight safety is well documented, hence effective interventions to further mitigate this problem are needed.

References

Field L. H. L., Edwards, S. D., Edwards, D. J., & Dean, S. E. (2018). Influence of HeartMath training programme on physiological and psychological variables. *Global Journal of Health Science, 10(2)*, 126. https://doi.org/10.5539/gjhs.v10n2p126

Henriques, G., Keffer, S., Abrahamson, C., & Jeanne Horst, S. (2011). Exploring the effectiveness of a computer-based heart rate variability biofeedback program in reducing anxiety in college students. *Applied Psychophysiology and Biofeedback*, *36*(2), 101-112. https://doi.org/10.1007/s10484-011-9151-4

Lee, E. (2012). Review of the psychometric evidence of the perceived stress scale. *Asian Nursing Research*, *6*(4), 121-127. https://doi.org/10.1016/j.anr.2012.08.004

McCraty, R. (2022). Following the rhythm of the heart: HeartMath institute's path to HRV biofeedback. *Applied Psychophysiology and Biofeedback*. https://doi.org/10.1007/s10484-022-09554-2

PRIME Road Markings for Motorcycle Casualty Reduction in Scotland

Prof Alex Stedmon[1], David McKenzie[2], Martin Langham[1], Kevin McKechnie[3], Richard Perry[2], Stuart Wilson[2], Morag Mackay[2], Stuart Geddes[3]

[1]Open Road Simulation Ltd, UK [2]Transport Scotland, UK [3]BEAR Scotland, UK

ABSTRACT

This paper presents research of new PRIME road markings in which over 32,000 motorcyclists were manually counted and coded. Analyses indicated that these unique road markings produced statistically significant reductions in motorcycle speed, improved lateral lane position, and reduced braking across 22 sites in the West Highlands of Scotland. This work provides insights into the 'Safe System' approach to support safer motorcycling and casualty reduction.

KEYWORDS

Motorcycle rider, Behaviour, Road markings, World trials, Casualty reduction

Introduction

Around the world motorcyclists are grossly over-represented in road traffic collision statistics (de Moraes, Godin, Dos Reis, Belloti and Bhandari, 2014; Transport Scotland, 2021).

Typically, motorcyclists are around 51 times more likely to be killed on the road than car drivers (Department for Transport, 2019, Transport Scotland 2020). These statistics highlight motorcyclists as one of the most vulnerable road user groups on public roads.

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 2021). In Scotland, motorcyclists represent only 2.2% of all registered vehicles but account for 14% of all Killed or Seriously Injured (KSI) casualties (Transport Scotland, 2020).

In many incidents, only the motorcyclist is involved and the 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 response, the Scottish Road Safety Framework has identified motorcyclists as a Priority Focus Area with a target for a 30% reduction in motorcyclists killed or seriously injured by 2030 (Transport Scotland, 2021).

PRIME road markings to support rider behaviour

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 driver might perceive and process risk factors in the environment around them (Gardener, Tate, Mackie, Stedmon, and Southey-Jones, 2017; Mulvihill, Candappa, and Corben, 2008).

From the motorcyclist's perspective, PCMs have been shown to influence rider behaviour in relation to speed, position, and braking to reinforce better rider behaviour (Hirsch, Moore, Stedmon, Mackie, and Scott, 2017; Hirsch, Scott, Mackie, Stedmon and Moore, 2018).

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 develop solutions that are cost effective to install and maintain. PRIMEs can be installed on existing roads quickly and efficiently or incorporated into road upgrade schemes.

'PRIMEs' provide a platform of innovative tools for motorcyclists with different riding styles. Motorcyclists are then able to adopt these tools and adapt their behaviour on approach to a potential hazard therefore optimising their expertise and enjoyment (and also their safety on the road). Of particular importance to this research programme was the safe navigation of bends. For this to occur, motorcyclists have to make sure that:

- speed – is suitable for the conditions
- position – is optimised for entering and travelling through the bend
- braking – is minimised whilst travelling around the bend

The PRIME road marking design investigated in this research comprised a series of three 'gateway' markings positioned on the approach to a bend. The intention was that the PRIME road marking would encourage motorcyclists to ride 'through the gap' and use the gateways as a cue to adjust their riding prior to the bend (Figure 1).

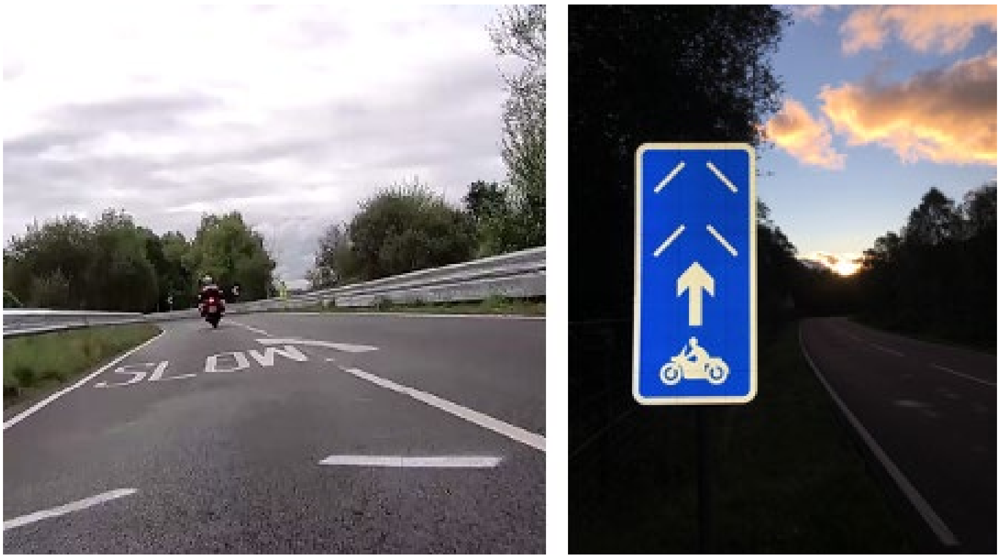


Figure 1: PRIMEs 'gateway' road marking and road sign

With a series of three PRIME gateway markings, there was potential for riders to adjust their braking point according to the motorcycle they were riding, their own riding style, or perhaps even due to weather and other environmental effects (i.e. in poor weather they might brake one marker back from their usual point).

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 review of collision cluster sites (BEAR, 2019, 2021). 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.

The trial sites were spread over a large geographic area of approximately 750 square miles ranging from Glencoe, Oban, Inveraray, Loch Lomond and towards Stirling and Crieff. They represented a range of bends on rural roads with speed limits over 40mph in line with recent casualty statistics (Transport Scotland, 2020).

Expert reviews were conducted for each potential site (e.g. complex geometry, tightening or double apexes, descents and inclines prior to bends, bends off fast sections of road) and 22 trial sites were identified and categorised as motorcycle cluster (MCL) sites or PRIME trial (PT) sites (Figure 2).

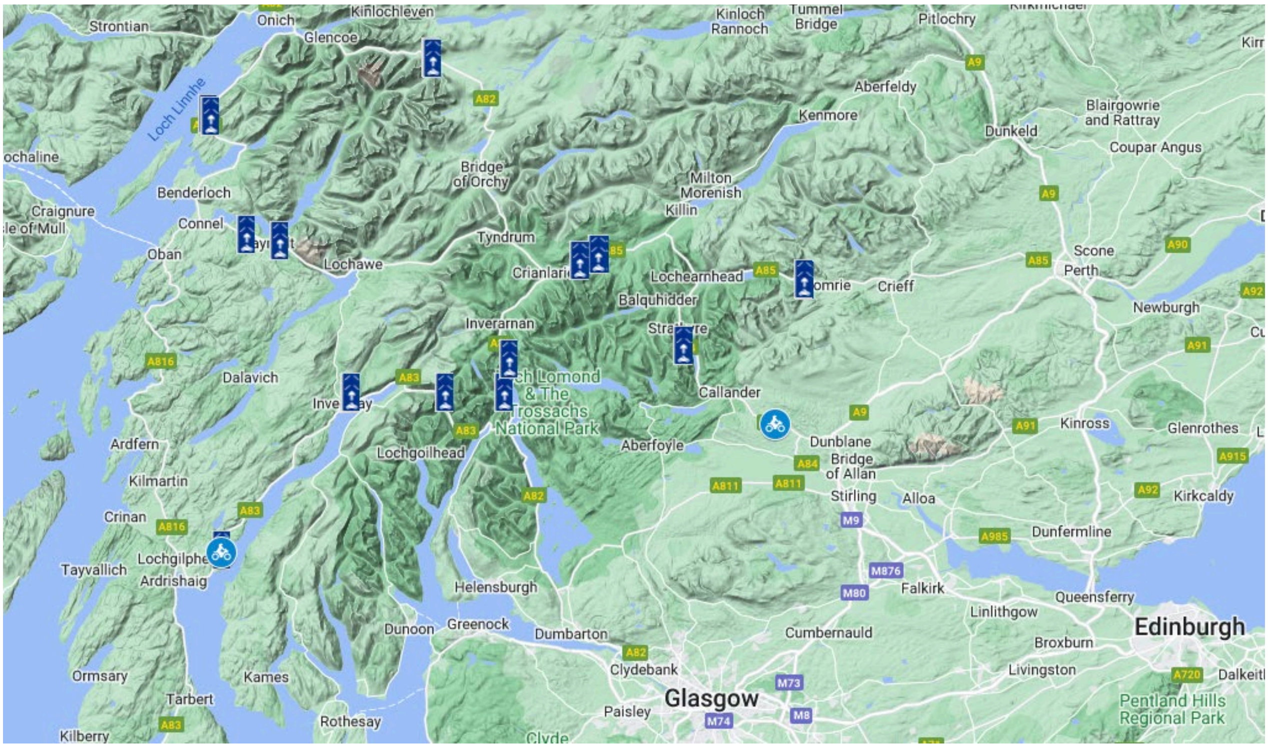


Figure 2: PRIMEs trial sites

Two comparison sites were also included where data were collected but PRIME road markings were not installed. Due to the wide variety of bends and road characteristics on the Trunk Road Network, these comparison sites were not regarded as experimental control conditions (i.e. where identical conditions are usually compared statistically).

Method

This research followed a conventional ‘pre- and post-intervention’ quasi-experimental paradigm, where 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 all the trial sites 32,213 motorcycles were observed and from these 9,919 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 in 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 100 mm wide 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
- PRIME 3 and 4 – to investigate the nature of sustained behaviour effects
- PRIME 5 and 6 – to investigate the nature of long-term behaviour effects

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.

In addition, rider interviews were conducted at the Green Welly Stop at Tyndrum and Inveraray waterfront as they were both popular meeting points and ride-out destinations for motorcyclists.

Procedure

Prior to data collection, trial sites were upgraded with various measures such as: resurfacing, line repainting, new crash barriers, vegetation removal, vehicle restraint systems (VRS), and motorcycle friendly 'bikeguard' installations. This meant that any extraneous variables were controlled as much as possible so that they would not otherwise influence rider behaviour (e.g. poor road surface, obscured views, potholes, poor safety provisions). Care was also taken to make sure that no changes to the sites were undertaken during the pilot 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 every trial site 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'). Where any difference was observed, Baseline 1 and Baseline 2 were kept as separate datasets and compared individually with the PRIME 1 and PRIME 2 datasets.

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.

Due to project restraints and as the data processing relied on specific and discrete manual counts, one researcher (Prof Stedmon) conducted the data processing and analyses. This researcher reviewed and re-checked data during the data processing activities. During the 2020 trials intra-rater reliability was assessed instead of inter-rater reliability (Stedmon, McKenzie, Langham, McKechnie, Perry and Wilson, 2021, 2022). 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 32,213 motorcycles were processed across all the trial sites and from these 9,919 lead motorcycles were analysed in more detail. Results from the 22 trial sites are summarised below (Table 1).

Table 1: Results for PRIME road markings across the 22 trial sites

Site	**Rider Behaviour**				
	Speed	**Position at PRIME**	**Position at Apex**	**Braking**	**Use of Gateway**
2020					
Appin House *north*		Sig			Trend
Appin House *south*	Sig	Sig	Sig	Trend	Sig
Kingshouse *north*	Trend	Sig			Sig
Kingshouse *south*		Sig	Sig	Trend	
Loch Lubhair *east*		Sig	Sig	Sig/Trend	Sig
Loch Lubhair *west*	Sig	Sig	Sig	Sig	Sig
Rob Roy's Dip *east 1*		Sig		Trend	Sig
Rob Roy's Dip *east 2*		Sig	Sig	Trend	Sig
Rob Roy's Dip *west 1*	Sig	Sig	Sig	Trend	Sig
Rob Roy's Dip *west 2*	Sig	Sig	Sig	Trend	Sig
2021					
Taynuilt	Sig		Sig	Trend	Sig
Inveruglas	Trend	Trend	Sig	Sig	Trend
Runacraig – *north*	Sig	Trend			Sig
Runacraig – *south*	Sig	Trend		Sig	Sig
Dunira	Sig	Sig	Sig	Sig/Trend	Sig
Bonawe	Sig		Trend	Trend	Trend
Landrick Bends	*no effect*	*no effect*	*no effect*	*Trend*	*no effect*
2022					
Dailnamac		Sig	Sig	Sig/Trend	Sig
Pulpit Rock			Sig	Sig/Trend	Sig
Butterbridge		Sig		Trend	Sig
Middle Kames	Trend	Sig		Sig	Sig
Salmon Draft – *north*	Sig			Sig/Trend	Sig
Salmon Draft – *south*	Trend	Sig	Sig		Sig
Carrick	*no effect*	*no effect*	*no effect*	*no effect*	*no effect*

These results are summarised below:

- **Speed** – statistically significant reductions in speed were observed at 10 trial sites. Trends were observed at four other sites
- **Lateral position at the final PRIME road marking** – statistically significant changes in lateral position were observed at 15 trial sites indicating that motorcyclists were riding in better positions on approach to the bend. Trends were observed at three other sites.
- **Lateral position at the apex of the bend** – statistically significant changes in lateral position were observed at 13 trial sites. A trend was observed at another site.
- **Braking behaviour** – statistically significant reductions in braking were observed at nine trial sites. Trends were observed at 15 other sites.
- **Use of the PRIME road markings** – statistically significant increases in the use of the road markings were observed at 18 trial sites. Trends were observed at three other sites.

At the comparison sites (indicated in *italics* in Table 1), as expected, no differences in rider behaviour were observed. At one site a trend for reduced braking was observed but this was due to temporary traffic management activity affecting traffic flow on specific data collections periods.

Rider interviews indicated that the majority of motorcyclists were supportive of PRIME road markings and felt they could be particularly useful to inexperienced riders or tourists with many riders stating that "anything that makes the roads safer is a good thing".

Discussion

Overall, the results for the PRIME road trials provide strong evidence for a range of beneficial effects of PRIMEs on rider behaviour on a range of bends. Across all three key measures (i.e. speed, position and braking) 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 PRIME 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′ outlining a long-term goal for road safety where no-one dies or is seriously injured by 2050 (Transport Scotland, 2021). It proposes a 'Safe Systems' 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 where roads have already been brought up to the best possible standard. 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 may 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 a 'Safe System' approach. There is clear evidence from the research conducted over the last 3-years that PRIMEs influence rider behaviour and it is important to begin planning for an implementation phase of work and address further research questions that will underpin the roll-out of PRIMEs more widely.

The project consortium have identified representatives from a Local Authority in Scotland with an interest in installing PRIMEs on their roads. This would provide an opportunity to widen the scope of PRIMEs in Scotland while also providing also ideal testbed for trialling a PRIMEs installation process (i.e. a user guide for authorities and councils so they can install PRIMEs without the need for expensive research).

Conclusion

This paper summarises a 3-year programme of PRIME road trials in Scotland funded by Transport Scotland and the Road Safety Trust. Throughout this work and the wider context of psychological theory, the approach taken has provided a planned and incremental development of understanding and building of evidence to take the work forward.

To date, 32,213 motorcycles have been manually counted and coded throughout the West Highlands with 9,919 lead motorcycles analysed in detail to understand the potential influence of PRIMEs on rider behaviour.

As far as the project consortium are aware, this makes the work the largest motorcycle behaviour investigation of its kind. Overall, the scientific evidence supports the concept that PRIMEs influence rider behaviour in positive ways by reducing speed, improving road position and reducing braking.

These findings underpin Transport Scotland's 'Road Safety Framework to 2030′ that has identified motorcyclists as a Priority Focus Area with a target of 30% reduction in motorcyclists killed or seriously injured by 2030 (Transport Scotland, 2021).

The concept of PRIME gateway markings provides a simple and very cost-effective solution to help reduce single vehicle crashes on our roads (which are one of the main collision types for motorcycles).

The evidence shows that if PRIMEs are installed they are used by motorcyclists and there have been no instances of a significant increase in speed, dangerous positioning, or increases in braking. These observations provide further evidence that PRIMEs did not have a detrimental effect on rider behaviour. In addition, since the start of the trials there have been no motorcycle injury collisions at any of the previously identified cluster sites.

The findings support the development of bespoke motorcycle road safety measures by Transport Scotland that provide an important step in reducing motorcyclist road casualties. By demonstrating the positive influence of PRIMEs on rider behaviour and rider safety, this work showcases Transport Scotland as a leader in this initiative for the UK and beyond.

References

BEAR Scotland Ltd. (2019). *North-West Unit Motorcycle Accident Investigation*. (unpublished report prepared for Transport Scotland, Trunk Road and Bus Operations (TBRO) Directorate. Scheme Ref: 18/NW/0801/042

BEAR Scotland Ltd. (2021). *2021 Motorcycle Accident Cluster Site and PRIME Site Selection Review*. (unpublished DRAFT report prepared for Transport Scotland, Trunk Road and Bus Operations (TBRO) Directorate. Scheme Ref: 20/NW/0801/062

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 (2019). *Reported Road Casualties in Great Britain: 2018 Annual Report*. The Stationery Office, London.

Department for Transport. (2021). *Vehicle Speed Compliance Statistics for Great Britain*. London: The Stationery Office.

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.

Hirsch, L., Moore, D., Stedmon, A.W., Mackie, H., and Scott, R. (2017). Keeping you in the loop: A human factors approach to motorcycling safety. In, *Proceedings of the Human Factors and Ergonomics Society of New Zealand Annual Conference*, 5-7 September 2017, Wellington, New Zealand

Hirsch, L., Scott, R., Mackie, H., Stedmon, A.W., and Moore, D. (2018). *Motorcycle safety on the Coromandel curves: The development and evaluation of perceptual counter-measures to influence rider speed, position and braking*. Prepared for the Accident Compensation Corporation

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., 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., 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

Transport Scotland (2020). *Reported road casualties Scotland 2019: Key findings report*. Transport Scotland, Glasgow: Crown Copyright.

Transport Scotland (2021). *Scotland's Road Safety Framework to 2030: Together making Scotland's roads safer*. Transport Scotland, Glasgow: Crown Copyright.

Human performance and automated operations: A regulatory perspective

Linn Iren Vestly Bergh[1], Kristian Solheim Teigen[1], Fredrik Dørum[1]

[1]Petroleum Safety Authority Norway

SUMMARY

The petroleum industry is becoming increasingly dependent on digital systems, and the companies have ambitious plans for increased use of digital technology – along the entire value chain. Increased levels of digitalisation present major opportunities for efficiency in the oil and gas industry and can also contribute to enhanced levels of resilience to major accident hazards. At the same time, new risks and uncertainties may be introduced. Based on developments in the industry and society in general, the Norwegian Petroleum Safety Authority (PSA) has in recent years pursued targeted knowledge development related to digitalisation and industrial cyber security. The PSA's follow-up activities related to digitalisation initiatives in the industry have been based on input and experience from several knowledge development projects. In this paper we will give insight into the main regulatory strategies we have used to follow-up initiatives in the industry, present results from audits on automated drilling operations and discuss the results from the follow-up activities in light of current regulatory development.

KEYWORDS

Automated operations, Automated drilling and well, Artificial intelligence (AI), Human-Automation interaction, Cyber security, Human performance.

Introduction

Norwegian Petroleum Safety Authority´s (PSA) goal is to follow-up that the petroleum activity gives high priority to safety, health and working environment when digital technology is developed, assessed, and implemented in the companies (PSA Dialogue, 2018; PSA, 2019).

The PSA has carried out several studies and research activities aimed at various aspects of digital technology and cyber security. Over 20 studies and knowledge reports relevant for the development and use of digital technology have been published on the PSA's webpages (Ptil.no). Studies and reports are developed in collaboration with external research. The findings are used as part of our prioritisation and planning of audits and follow-up.

One of PSA's main areas of concern is related to how increased automatization effect human performance in drilling operations. Based on results from studies and developments in the industry, the PSA has in recent years initiated several supervisory activities targeting automated drilling and well operations.

Automated systems and human performance

The level of automatization in operations ranges from systems, where personnel have overall control over most operations, to systems that work completely independent from human intervention (Johnsen et al. 2020; Kaber, 2018). Despite increased automation, the industry will in many cases use systems where personnel have an important role in monitoring them. If an unforeseen situation arise, personnel must evaluate and control a complex situation without having sufficient time, knowledge or overview (Ottermo et al. 2021; Johnsen et al. 2020).

Companies in the petroleum sector implement more advanced digital technologies such as artificial intelligence (AI) and machine learning (ML). Increased use of automated systems may introduce new types of risks and vulnerabilities (Johnsen et al. 2020; Endsley, 2019; Endsley, 2023). Most near misses and incidents involving human automation operations arise from a mismatch between the properties of the system as a whole and the characteristics of human information processing (Endsley, 2019; Endsley, 2023). Development within digital technologies have altered the human-computer interface. On one hand, it has contributed to a reduction in manual and physical tasks. On the other hand, it has changed the demands with regards to cognitive processing (Johnsen, 2020; Ernstsen, 2021; Longo et al., 2022). According to research within human factors engineering it is essential to incorporate a strong focus on how humans use digital technology in a safe way. To do this, knowledge on human cognition should be included in early technology development (Johnsen, 2020; Ernstsen, 2021).

Several researchers argue that effective digital systems rely on selection of a data model which has a line of reasoning that explains it's behaviour – thereby optimising human performance. As such, it is important to consider methods for evaluating the user interaction and interpretation of the data model (Endsley, 2023; Bansal, 2019). A human-centred design can mitigate high mental workload, lack of situational awareness, alienation, knowledge degradation and fatigue that may negatively affect people's ability to monitor and intervene when needed. This will not only strengthen safety but also make the operation more reliable and efficient (Johnson, 2020; Ernstsen, 2021; Endsley, 2019).

Safe and effective interaction between human and technology is important for ensuring safety. Important factors in this regard are mental models, transparent computer models, trust in technology, and function allocation (Ernstsen, 2021). Literature and development of EU artificial intelligence (AI) regulations points to the importance of developing digital systems that are human-centred. Interesting areas of research have developed in line with the advancements in digital technology. One of the areas receiving increased attention is research within explainable artificial intelligence (XAI) (The Royal Society, 2019). As advanced digital solutions such as AI technologies become embedded in decision-making processes it will be important to ensure individuals developing AI, or subject to an AI-supported decision, understand how the system works. AI solutions applied today can produce precise results, however their reasoning is also highly complex. AI models that are so complicated that experts cannot fully understand them are called black-box. As such, XAI involves developing solutions where the human operator can interpret and understand why a system takes certain actions, decisions or makes predictions. (The Royal Society, 2019; European Commission, 2019). As researchers continually try to develop more transparent digital solutions there has also been an increased attention on the need for XAI to draw on insights from social sciences. For example, Miller (2019) argues that XAI should ensure knowledge about how humans' natural way of presenting and evaluating information are included when advanced data models are developed.

Another interesting area of research is the Human-centred artificial intelligence (HCAI) framework. This framework is focused on system design and the development of reliable, safe and trustworthy systems in safety-critical operations. Ensuring both high levels of human control and high levels of computer automation to increase human performance are highlighted as critical topics (Shneiderman, 2020). As digital technology becomes more complex, an important area for future automation is system design that is flexible and adaptive to the current status of the operator, such as stress level, fatigue and level of attention (Johnson, 2020).

Automated systems in drilling and well

In the last decade automatization has been a key driver for increased drilling performance, reduced well cost and improved safe well delivery. Several automated solutions have been developed and implemented, gradually changing work tasks and processes from manual operations of machines to automated solutions. For example, digital technology is increasingly being applied to support the driller in analysing, interpreting, and making decisions for further actions. Technologies in automated drilling solutions can include offline and realtime models such as; digital twins of the wellbore and geology, simultaneous multi machine control, physical models of wellbore mechanics, automated fluid handling and well control. Each area has varying degrees of autonomy (Ottermo et al. 2019).

On the Norwegian continental shelf Equinor first tested Automated Drilling Controls (ADC) together with Transocean in 2017 (Offshore Technology, 2019), and has since expanded its use of the technology to the majority of its contracted mobile drilling units. The main benefits of drilling automation are reduced overall cost, consistency of operations through reductions of errors and reduction of people required on board (Hereira, 2021). Over the last years there has also been some interesting research developments related to autonomous drilling. The world's first autonomous drilling was demonstrated in 2021 by a research group from the Norwegian Research Centre (NORCE). The autonomous drilling system was tested both in a virtual environment and at a test rig, called Ullrigg (Mihai et al., 2022).

The Drilling Automation Roadmap, a joint industry project backed by the SPE Drilling Systems Automation Technical Section (DSATS) describes that automation enables drilling of more challenging wells and drilling through formations that has not previously been possible. Drilling for hydrocarbons is a high-risk operation, involving high pressures, heavy equipment, and operations in harsh environment. Thus, errors and accidents can have enormous consequences for humans, the environment and the organisations and equipment involved. With increased use of robotics and remote control, human presences and exposure to hazards can be reduced. However, the consequences for environment remain.

In the following we will give insight into PSA's follow-up initiatives in the industry and discuss the results from the follow-up activities.

Methods

As a regulator, the PSA use different methods and approaches in our follow-up of the automatization initiatives. Within drilling and well operations, we have executed several audits related to automated operations and human performance over the past years. The PSA ideally follow digitalisation projects from early design phase to testing / qualification, building and implementation. To understand development of overall risk, the PSA assess whether the companies are pursuing their operations prudently and in accordance with the regulations. The PSA also see to that the companies actively promote Health, Safety and Environment (HSE) when digital solutions are implemented.

This paper explores the findings from three audits. The objective of these audits were to assess whether companies prioritized safety and human factors when digital technologies were deployed and implemented. This included issues related to human performance, compliance with regulatory requirements for the implementation and use of automated drilling operations, robotisation of pipe handling and digital well planning. The audits were aimed at drilling and well operators, rig owners and related service companies.

The auditing teams from PSA were multidisciplinary, ensuring a holistic human, technology, organisation perspective. The audits were carried out as a combination of document reviews, meetings, semi-structured interviews with operating companies, drilling contractors and service companies as well as field observations conducted onshore and offshore. Both operating personnel and management were interviewed. Information collected through the document reviews, interviews and field observations from the audits were structured and analysed. Where the observations constituted lack of compliance to regulations, non-conformances were issued. Common issues and findings from these three audits were systematically assessed and categorised into main topics, forming the basis for the results presented in this study.

27 interviews and 10 group sessions and meetings were performed. Examples of topics that were highlighted in interviews were: Technology qualification basis, risk and technology assessments, organisational analysis, and other requirements and acceptance criteria for safe development and implementation. Other topics included how new technology was applied in the drilling and well operations, impacts on HSE effects and how risk was handled if or when the technology failed. We also assessed how implementation of new systems impacted work assignments, tasks, and processes. Further how personnel had been trained, and prepared for changes in technology, organisation, and work execution.

Several regulatory requirements in the Norwegian petroleum Health, Safety and Environment (HSE) regulations are relevant when following-up companies with regards to the development and use of automated drilling. The regulatory requirements listed in below table are relevant for technology development, ergonomic design, the interface between human and computer as well as control and monitoring systems.

Table 1. Relevant regulations when auditing companies in the industry

Regulations	Section in regulations
The Management Regulations	Section 16 General requirements regarding analyses Section 18 Analysis of the working environment
The Facilities Regulations	Section 9 Qualification and use of new technology and new methods Section 10 Installations, systems and equipment Section 20 Ergonomic design Section 21 Human machine interface and information presentation Section 34a Control and monitoring systems
The Activities Regulations	Section 21 Competence Section 23 Training and drills Section 24 Procedures
The Technical and Operational Regulations	Section 21 Human-machine interface and information presentation Section 33a Control and monitoring system

The management regulations stipulate that the industry bears the responsibility to actively prevent harm or danger of harm to people, the environment or material assets in accordance with the HSE legislation. This includes internal requirements and acceptance criteria that are of significance for complying with requirements in the regulation. In addition, the risks shall be further reduced to the lowest extent possible. The regulation requires the companies to manages all risks when implementing new digital technology. It also stipulates that technical, organisational solutions must be developed in a Human, Technology and Organisational (HTO) perspective.

In the next section the results from the audits have been categorised into six main topics.

Results

Technology development and technology qualification

Limited attention is paid to human abilities and prerequisites when developing digital technology. Human abilities and prerequisites are often not considered or included in the technology development process (PSA, 2018; PSA, 2021; PSA, 2022). The development of digital solutions was outsourced to subcontractors and / or digital solutions were bought "off the shelf" and retrofitted to existing systems. This was done without sufficiently ensuring that the technology was appropriately qualified for its intended use, which resulted in "safety risks that flew under the radar" (PSA, 2022). Even though new digital solutions often required limited technical installation, our findings show that they pose significant implications for operations and human performance, and thus require due assessments.

Increased complexity and interfaces

Digital technology and drilling automation solutions are complex and can be difficult to understand for personnel. Functionality is coded and integrated in the software, and thus hidden from users, as well as those who perform risk assessments of operations. Although some of the individual digital technologies implemented could be perceived as relatively understandable, the sum of these - and how they interacted with each other – were less transparent to the user (PSA, 2021; PSA, 2022). We found non-conformances concerning lack of integration between the controls systems, and deficient integration between different user interfaces in the driller's cabin. Weaknesses in the Human Machin Interface (HMI), high alarm rates, and ongoing installation, troubleshooting and implementation of upgrades and software changes while the rig was in full operation contributed to fatigue and stress. Further the increased complexity contributed to reduced situational awareness, both for the operator and other partakers in the drilling operation (PSA, 2022). Operating personnel also expressed a sense of insecurity, and fatigue related to scale and pace of change. It emerged that for most workers in the companies where a high degree of automation was introduced, there were also several changes and digitalisation initiatives impacting other areas of their workday (PSA, 2021; PSA, 2022).

Measuring performance and learning

Technology provides an increased opportunity to measure performance. However, we found that the focus, scope and level of reporting contributed to increased time pressure and negatively impacted human performance. The use of Key Performance Indicators (KPI) were mainly geared towards efficiency and speed. For example, it was common to use KPIs and micro KPIs where individual work tasks and operations were measured in minutes and seconds. Status and progress on individual KPIs were presented in daily meetings. However, the companies were unable to show how learning from the KPIs were used to improve the technology, or in other ways contribute towards overall risk reduction (PSA, 2021; PSA, 2022).

Legal requirements and standards

There was a lack of knowledge and clarity with regards to interpreting regulatory requirements in a human-centred design approach. Therefore, in some of the cases, relevant subject matter experts had not been included at an early stage in the technology development. Furthermore, there seemed to be a lack of understanding of how the functional requirements in the HSE regulations were applicable in the AI domain (PSA, 2022; PSA, 2018). The PSA experience that the industry calls for standards and methods for a human-centred approach to digital technologies. Moreover, there also seems to be a lack of industry understanding of how functional requirements in existing regulation may also apply for digital solutions.

Aligning work processes and technology

The audited parties clearly emphasized that increased use of digital technology offshore is a prerequisite for successful introduction of new ways of working. These changes to ways of working may help to simplify and improve decision support for the personnel involved, but also leads to changes in roles and responsibilities and introduces new competence requirements. For example, the primary work task for a driller changes from manually adjusting drill bit rotation and fluid flow, to monitoring and being ready to intervene if the automated drilling process fails or needs adjustments. However, we found that the audited parties had challenges with succeeding in adapting and changing work processes at the same time as the automated solutions were introduced. This resulted in a mismatch between the technology and the work processes. Further we found that the audited companies had not established routines, or defined in the written work tasks, how long at a time executing personnel should remain in the operator's chair. Even when loss of situational awareness for the operator was identified as a risk, we found that evaluation and mitigating actions were neither identified nor implemented. Several non-conformances were identified concerning procedures and work processes, including high workload, lack of training and role unclarity (PSA, 2021; PSA, 2022).

Risk assessments in operations

Risk assessments conducted for addressing local operational risk factors, when introducing new technology and ways of working were deficient. For example, there was an expectation that the automated mode was the default for operations. However, interviewees reported that the threshold for assuming manual control or choosing to conduct operations manually was high, and not clearly defined in procedures, risk assessments and risk registers. Risk factors related to changes in mode were not identified or evaluated. Operating personnel were often unaware of risks related to changes in mode between manual and automated operations (PSA, 2021; PSA, 2022).

Discussion

This paper has presented results from the PSA's audits and follow-up of automated operations and human performance in automated drilling technology and operation.

Results from PSA's audits and studies show that digital systems are complex and can be difficult to understand for operating personnel (Johnsen, et al. 2020; Ottermo, et al. 2019, Erntsen et al. 2021; Gressgård et.al 2018; PSA, 2021; PSA, 2022; PSA 2018). AI and digital models which are coded and integrated into a software are less transparent to the operator using the technology. This implies that human performance in digital systems relies on a system that can convey the systems actions in a transparent manner. Lack of transparency and explainability in the interface can lead to operators experiencing inability to interpret information and predict system behaviour and automated action (Endsley, 2023). Understanding and being able to predict drivers for human performance is a key issue in the context of safety-critical behaviour and designing technology that can mitigate undesirable mental states (Endsley, 2023; Roberts et al., 2015). Factors such fatigue, distractions, and stress, can have an adverse impact on operators working memory and task performance (Johnsen et al. 2020). In a high-risk industry, such as the petroleum industry, minute decisions can have severe consequences. Thus, managing these risks are critical for achieving a prudent level of safety, and a priority for the PSA.

Researchers argue that complexity, involvement and workload affect the human-automation interaction (Johnsen et al., 2020; Endsley, 2019). Therefore, it is necessary to examine how new work tasks are designed. Further how the design of new work process takes human's strengths and limitations into account. This is in line with findings from the PSAs audits. Introduction of automated drilling systems in some cases created a perceived distance between the operators and the risk factors associated with the work task, e.g. well control. The operators found that they to a larger degree were in a pacified state, supervising and monitoring the system conducting the operations – as opposed to actively taking part in the operation. Some described it as they were losing their "feel for" the well, and that it could be difficult to maintain good situational awareness that included well control factors. This is in line with research showing that automated systems can lead to impaired mental models and reduced situational awareness (Ottermo et al., 2020). Another area of concern in the literature is that automation may increase the overall mental strain on the operator (Johnsen et al. 2020). Our findings further showed that increased automation could also be a driver for prolonged sessions in the operator's chair, and that this imposed a mental strain, reducing operator's vigilance and sense of situational awareness.

The audited companies that were early adopters of technology, tended to digitalize across business functions and utilities. Meaning that the same worker whose primary work tasks were affected by introduction of automation, was also exposed to several other new digital solutions in other and remaining work tasks (PSA, 2021). As a regulator we are concerned that the sum of changes contribute to a state of digital fatigue, even if none of the changes or systems by itself can be considered overwhelming. Systems that are primarily seen as safe, still challenge the organisational boundaries and practices when autonomous systems are introduced (Oliver et al. 2017; Johnsen et al. 2020). Problems with the automated system often occur in unknown and unexpected situations. Reports highlights that this must be dealt with at a human-automation level as well as on an organisational level, which includes considerations of interaction between many stakeholders (Johnsen et al., 2020; Endsley, 2017; Gressgård et al., 2018). Training, trust in the change processes and the digital technology are organisational factors that must be followed up when introducing automated and autonomous systems. The importance of having a lifecycle perspective (e.g. assessing and evaluating risk in early development and in operations) is critical with regards to preventing negative impact as well as optimizing opportunities that digital technology provides (NIST, 2023).

In a technology-intensive industry we have found there is an uncertainty with regards to how the regulatory requirements should be understood. This is not isolated to the petroleum industry (Gressgård, 2018; EU Commission). EU efforts are also being prioritised on ensuring sufficient regulations for digital technologies such as AI. The newly proposed EU regulations require development and use of AI to be human-centred, trustworthy, and based on ethical principles. The EU proposes a risk-based approach to division into risk categories. This means strict regulation of so-called high-risk AI systems. Furthermore, risk management throughout the life cycle of the system for AI technology must be established, including requirements for development, testing, evaluation, and implementation of risk-reducing measures (EU Commission). In 2020, the Norwegian government published a national strategy for AI. In line with the EU regulations this strategy highlights the need to tackle potential challenges such as data quality, transparency and autonomy when developing and using AI. In recent years recommended practises for performing verification and assurance activities for data driven algorithms (DNV-RP-510) management of risk in AI (NIST-AI100-1), and qualification of digital twins (DNV-RP-A204) have emerged. However, as these are relatively new contributions in the standardization domain, their informative effects on regulators and sectorial directives and regulation are yet to materialize.

Although the PSA consider its current regulation relevant within digital domain, we continually evaluate the regulations applicability. Furthermore, it is important that our regulation refer to the appropriate and relevant standards in relation to digital systems. It is expected that norms and standards are at the forefront of the digital development. This is an industry responsibility, and a part of what the authorities must assess in relation to regulatory development. In the way forward, there is a need for to assess and conclude what AI means in terms of regulatory activities and regulations. For the PSA it is important to contribute to the safe use of advanced digital technology such as AI in high-risk context. An important part of this is ensuring up-to date requirements for the development, deployment, and use of AI systems. This may include requirements covering transparency, explainability, and assessments of systems, as well as requirements for cyber security. The PSA will continue to assess whether the companies are pursuing their digital endeavours prudently, in accordance with the regulations and actively promoting Health, Safety and Environment (HSE) when digital solutions are implemented.

Future directions:

Going forward it will be relevant to evaluate how knowledge on human performance apply to the cyber security domain. Digitalisation and automatization of real-world physical assets may impact achieved level of safety (Nelson et al., 2021). Automated systems require a high level of availability, connectivity and thus web exposure through network interfaces. A system that is accessible through the web, is inherently more vulnerable to attacks where unauthorised persons access sensitive information or target critical functions from anywhere in the world (Ottermo et al, 2019). As such digitalisation of industrial assets come with a substantial increase in security risk (Rubio et al.2019). Combating this risk entails collection and monitoring of large-scale data sets and logs over network traffic. To detect anomalous traffic and intrusions, AI and machine learning technology is increasingly applied in intrusion detection systems (IDS) (Lee et al, 2022; Rubio et al, 2019). In the same manner as AI systems within drilling and well, AI enabled IDS systems are inherently complex and challenging for human operators to understand. As Lee & co argues *"analysts have little choice but to trust the AI-predicted outcomes. In the field, even a security control center with a high-performing IDS system eventually requires validation by a human analyst.*"(Lee et al. 2022). Thus, there is a need for transparency to support optimal human performance, also in the cyber security domain.

Conclusion

As PSA's follow-up and this study has shown, the petroleum industry is becoming increasingly dependent on digital systems. As digital technology is taking over manual tasks, employees still play an important role for safety in the sector. Increased levels of digitalisation present major opportunities for efficiency and can also contribute to enhanced levels of resilience to major accident hazards. At the same time, new risks and uncertainties may be introduced. This means that several technical, organisational, and human challenges must be systematically followed up to realise the potential offered by digital technology. In the way forward, the PSA will contribute to promote safe use of digital technology. An important part of this is ensuring up-to date requirements for the development, deployment, and use of digital systems.

The companies are responsible for safe operations. Therefore, they must assess vulnerability and risk from an integrated perspective which includes human, technological and organizational (HTO) aspects. Each company must take ownership of and manage the risk related to the implementation of new systems and technological solutions.

References

Bansal, G., Nushi, B., Kamar, E., Lasecki, W. S., Weld, D. S., & Horvitz, E. (2019). Beyond Accuracy: The Role of Mental Models in Human-AI Team Performance. *Proceedings of the AAAI* Conference on Human Computation and Crowdsourcing, *7*(1), 2-11.

DNV, (2020). DNV RP-A204, Qualification and assurance of digital twins. https://www.dnv.com/oilgas/download/dnv-rp-a204-qualification-and-assurance-of-digital-twins.html

DNV, (2020). DNV RP 0510, Framework for assurance of data driven algorithms and models. https://www.dnv.com/software/campaigns-2020/dnvs-recommended-practice-for-machine-learning-assurance-rp.html

Endsley, M.R. (2019). Human Factors & Aviation Safety, Testimony to the United States House of Representatives Hearing on Boeing 737-Max8 Crashes — December 11, 2019

Endsley, M. R. (2017). From here to autonomy: Lessons learned from human- automation research. Human Factors, 59(1), 5–27.

Endsley, M. R. (2023). Supporting Human-AI Teams: Transparency, explainability, and situation awareness. Computers in Human Behavior, 140, 1-15.

Ernstsen, J., Aulie, E. G., Sætren, G. B., Stenhammer, H. C., & Phillips, R. (2021). Et menneskesentrert perspektiv på kognitiv teknologi i petroleumsindustrien. 2021. Safetec. Petroleumstilsynet. Available on: https://www.ptil.no/globalassets/fagstoff/prosjektrapporter/ikt-sikkerhet/digitalisering-mto-safetec-st-16407-2-hovedrapport-rev.-3-1.pdf

EU Commission. Shaping Europe's digital future. Regulatory framework proposal on artificial intelligence. Accessed at: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Gressgård, L. J., Melberg, K., Risdal, M., Selvik, J. T., & Skotnes, R. Ø. (2018). Digitalisering i petroleumsnæringen.IRIS. Petroleumstilsynet. Avaliable on: https://www.ptil.no/globalassets/fagstoff/prosjektrapporter/ikt-sikkerhet/digitalisering-i-petroleumsnaringen-1.pdf

Guzman, N.H.C., Kozine, I., Lundteigen, M.A. (2021). An integrated safety and security analysis for cyber-physical harm scenarios. Safety Science. 144.

Heredia, J., Tengesdal, J. E., Hobberstad, R., Marck, J., Kleivenes, H., and Marcel E. N. (2021). Drilling Automation – Benefits of the New Drilling Model. Paper presented at the Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE.

Johnsen, S.O., Holen, S., Aalberg, A.L., Bjørkevoll, K.S., Evjemo, T.E., Johansen, G., Myklebust, T., Okstad, E.,Pavlov, A., & Porathe, T. (2020). Automatisering og autonome systemer: Menneskesentrert design. Petroleumstilsynet. Avaliable on: https://www.ptil.no/globalassets/fagstoff/prosjektrapporter/ikt-sikkerhet/prosjektrapport-sintef-automatisering-og-autonome-systemer-menneskesentrert-design.pdf

Lee, y., Lee, E. Lee, T. (2022). Human-centered efficient explanation on intrusion detection prediction. Electronics, 11, 2082. Available on: https://www.researchgate.net/publication/361713670_Human-Centered_Efficient_Explanation_on_Intrusion_Detection_Prediction

Kaber, D. B. (2018). Issues in human–automation interaction modeling: Presumptive aspects of frameworks of types and levels of automation. Journal of Cognitive Engineering and Decision Making, 12(1), 7-24.

Mihai, R., Cayeux, E., Daireaux, B., Carlsen, L., Ambrus, A., Simensen, P., Welmer, M., and Matthew J. (2022). Demonstration of Autonomous Drilling on a Full-Scale Test Rig. Paper presented at the SPE Annual Technical Conference and Exhibition, Houston, Texas, USA.

Miller, T. (2019). Explanation in artificial intelligence: Insights from social sciences. Artificial intelligence. 267, 1-38.

NIST (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). Available on: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

Offshore Technology (2019). Transocean and Equinor sign agreement to install ADC systems on rigs. Available on: https://www.offshore-technology.com/news/transocean-equinor-adc-systems-rigs/

Oliver N., Calvard, T., Potočnik, K. (2017). Cognition, Technology, and Organizational Limits: Lessons from the Air France 447 Disaster. Organization Science 28(4):729-743.

Ottermo, M.V., Wille, E., Bjørkevoll, K.S., Bodsberg, L., Evjemo, T.E., Fjørtoft, K., Jaatun, M.G., Myklebust, T., Okstad, E. (2021). Rapportering av hendelser i automatiserte systemer I boreoperasjoner. Petroleumstilsynet. Avaliable on: https://www.ptil.no/globalassets/fagstoff/prosjektrapporter/ikt-sikkerhet/digitalisering-mto-sintef-sluttrapport-ptil-varsling-av-hendelser-2021-12-16---signert.pdf

Petroleum Safety Authority (PSA) (2022). Equinor/KCA Deutag – Oseberg South/Askepott – technology development, implementation and use of digital well planning and automated drilling control. Available on: https://www.ptil.no/en/supervision/audit-reports/2022/equinorkca-deutag--oseberg-southaskepott--technology-development-implementation-and-use-of-digital-well-planning-and-automated-drilling-control/

Petroleum Safety Authority Norway (PSA) (2021). Equinor and Transocean – well planning and automated drilling control – Transocean Enabler. Available on: https://www.ptil.no/en/supervision/audit-reports/2021/equinor-og-transocean-well-planning-and-automated-drilling-control-transocean-enabler/

Petroleum Safety Authority Norway (PSA) (2018). Rapport etter tilsyn med planlagt ombygging, digitalisering og robotisering av boreanlegget på Valhall IP - Performinator-prosjektet. Available on: https://www.ptil.no/contentassets/d059c982c03e46c397d629b5606c0acf/2020_300-revidert-rapport-etter-tilsyn-med-planlagt-ombygging-digitalisering-og-robotisering-av-boreanlegget-pa-valhall-ip---performinator-prosjektet-1303202000000.pdf

Petroleum Safety Authority Norway (PSA) (2022). IKT-sikkerhet i industrielle systemer. Available on: https://www.ptil.no/fagstoff/utforsk-fagstoff/fagartikler/2021/ikt-sikkerhet-i-industrielle-systemer/

Petroleum Safety Authority Norway (PSA) (2022). New technology must take people into account. Available on: https://www.ptil.no/en/technical-competence/explore-technical-subjects/news/2022/new-technology-must-take-people-into-account/

Roberts, R., Flin, R., Cleland, J. (2015). "Everything was fine": An analysis of the drill crew's situation awareness on Deepwater Horizon. Journal Loss Prevention Process Ind. 38, 87–100.

Rubio, J.E., Alcaraz, C., Roman, R., Lopez, J. (2019). Current cyber-defense trends in industrial control systems. Computers & Security, 87.

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. International Journal of Human–Computer Interaction, 36(6), 495-504.

The Drilling Systems Automation (DSA) Roadmap Report (2022). Available on: https://dsaroadmap.org/drilling-systems-automation/dsa-r-report/

CRIT-UK: A tool to understand contributory factors involved in current cyclist incidents

Siobhan E. Merriman[1], Katie J. Parnell[1] & Katherine L. Plant[1]

[1]Human Factors Engineering, Transportation Research Group, Faculty of Engineering and Physical Sciences, University of Southampton, UK.

SUMMARY

Collisions between motorised vehicles and cyclists remain a persistent road safety issue worldwide, however the nature of these collisions remain poorly understood as there are currently few mechanisms available for cyclists to report sufficient detail about their collisions and near-miss incidents. Originally developed in Australia, this paper will describe the expansion of the Cyclists Report of Incidents Tool (CRIT) app to the UK road system, to understand the contributory factors involved in current cyclist collisions and near-miss incidents in the UK. Furthermore, it will explore how these factors may change with the introduction of Levels 2 and 3 Automated Vehicles.

KEYWORDS

Cycling, Collisions, Near-Miss Incidents, Contributory Factors, Automated Vehicles

Introduction

Private vehicle travel is one of the biggest contributors of greenhouse gas emissions worldwide, including in the UK (Department for Transport (DfT), 2021a). As concerns over climate change grow, there is a need to shift towards more sustainable and active transport modes such as cycling. However, collisions between cars and cyclists remain a persistent road safety issue worldwide. Despite the greater number of cars on the road, in the UK in 2020, cyclists had a higher fatality rate per billion vehicle miles compared to car occupants (27 vs 3: DfT, 2021b). This could deter people from cycling, which may prevent the sustainability and health benefits that cycling offers from being realised. Additionally, with the introduction of Levels 2 and 3 Automated Vehicles (AVs) which have systems that control some or all of the driving tasks (Society for Automotive Engineers, 2021), the interactions between cyclists and motorised vehicles will change, so the characteristics of these incidents may change. As such, an understanding of the contributory factors involved in current cyclist incidents is needed in order to develop interventions to prevent them and gain an understanding of how these incidents may change with the introduction of Levels 2 and 3 AVs.

A recognised approach for enhancing our understanding and prevention of incidents is the use of an appropriate incident reporting and learning system (Goode, et al., 2018). However, there are few mechanisms available for cyclists to report sufficient detail about their collisions and near-miss incidents. For example, the STATS19 system is used in the UK to record road accident details. However, this is completed by the police and does not capture near-miss or no injury incidents. Furthermore, it only focusses on the immediate surroundings, vehicles and road users involved, rather than taking a broader systems perspective to capture higher level factors e.g. local councils, legislation and government (McIlroy, et al., 2021). This paper will describe the expansion of the Cyclists Report of Incidents Tool (CRIT) app originally developed in Australia, to the UK (CRIT-UK). This tool enables the reporting and analysis of cyclist collisions and near-miss incidents, which may enhance our understanding of the contributory factors involved and inform the development of interventions to enhance future cycling safety. Regulatory and government agencies, such as the DfT, will be able to use the findings to make evidence-based decisions and recommendations (e.g. infrastructure improvements, law changes) based on real, as opposed to the perceived risks, associated with cyclist collisions and near-miss incidents. This work will describe the app, present initial findings from a UK trial and explore how these incidents may change with the introduction of Levels 2 and 3 AVs.

CRIT-UK App

The CRIT-UK app can be downloaded onto a smart phone device. When a collision or near-miss incident occurs, cyclists are asked to report the incident on the app. The app records the date and time of the incident, form of activity (on-road, off-road), type of incident (collision or near-miss), location, an incident description, contributory factors involved (selected from a list), severity and treatment (none, immediate first-aid, hospital). The list of contributory factors allows cyclists to select high-level factors (e.g. other road users, environment, equipment, cyclist, local councils) and sub-factors (e.g. driver behaviour, road rules, surface, obstacles and debris, cycling infrastructure) which they believe contributed to the incident. Once a week, cyclists are also asked to report the number of hours that they have cycled in the past week, to enable the calculation of incident rates.

The data is self-reported and is not being correlated to more objective data sources (e.g. CCTV, dashcam footage), so there is an element of subjectivity and bias which may influence the accuracy of the reports and the recommendations that are made. The CRIT-UK app is reliant on cyclists being able to remember and accurately report all contributory factors involved, it does not include fatal incidents and only analyses the incidents which have been reported on the app. However, a mechanism to report fatal incidents does exist (e.g. STATS19, see above), so the findings from the CRIT-UK app relating to collisions and near-miss incidents can be compared to these fatal incidents to highlight where interventions are needed. The details are coming from the cyclists themselves and their perceptions of safety, rather than interpreted by a third party. Aggregated analyses from multiple incidents over time can highlight the most frequently reported contributory factors which should be given the greatest focus in future interventions to prevent similar incidents from occurring.

Preliminary Findings from Australia

A six-month trial has been conducted in Australia (Cox, et al., 2022). Between December 2021 and June 2022, 316 cyclists (248 males) used the app and 109 incidents were reported (92 near misses). Most incidents occurred on the road (92%) and in the morning (73%) and were perceived to be minor (44% for collisions, 55% for near misses). For both types of incidents, cyclists frequently reported the driver’s behaviour as a contributory factor, with the sub-factors of “pulling out in front of cyclist” and “non-compliance with road rules” (near-miss only) being the most frequently reported. The road infrastructure (roundabouts, intersection/junction) and cyclist infrastructure (bicycle lane and lack of bicycle lane) were frequently reported contributory factors. Cyclists also reported factors beyond the equipment, road environment and road users including factors relating to vehicle and infrastructure maintenance and repairs, driver education and training and media and social media. These initial findings demonstrate that useful data is being gathered about the contributory factors involved in current cyclist collisions and near-miss incidents in Australia. This app is being deployed in the UK in early 2023, and this work will present initial findings from this trial and consider how these incidents may change with the introduction of Levels 2 and 3 AVs.

Conclusion

The CRIT-UK app allows cyclists to report collisions and near-miss incidents. This can enhance our understanding of the contributory factors involved and inform the development of interventions to enhance cycling safety now and when Levels 2 and 3 AVs are introduced into the road network.

References

Cox, J., McLean, S., Hulme, A., Read, G., & Salmon, P. (2022). *The Cyclist Reporting of Incidents Tool. Australian National Incident Dataset for Cyclist Incidents.* https://www.usc.edu.au/research/centre-for-human-factors-and-sociotechnical-systems/the-cyclist-reporting-of-incidents-tool-crit [Accessed 23/11/2022].

Department for Transport. (2021a). Transport and Environment Statistics 2021 Annual Report. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/984685/transport-and-environment-statistics-2021.pdf [Accessed 23/11/2022].

Department for Transport. (2021b). *Reported road casualties Great Britain, annual report:2020.* https://www.gov.uk/government/statistics/reported-road-casualties-great-britain-annual-report-2020/reported-road-casualties-great-britain-annual-report-2020#headline-figures [Accessed 23/11/2022].

Goode, N., Salmon, P. M., Lenné, M. G., & Finch, C. F. (2018). *Translating systems thinking into practice: a guide to developing incident reporting systems.* CRC Press.

McIlroy, R. C., Plant, K. L., & Stanton, N. A. (2021). Intuition, the Accimap, and the question "why?" Identifying and classifying higher-order factors contributing to road traffic collisions. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 31, 546–558.

Society for Automotive Engineers. (2021). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles: J3016_202104.* https://saemobilus.sae.org/content/j3016_202104 [Accessed 23/11/2022].

Situation Awareness in Midwifery Practice

Rachael L. Martin[1] & Paul Bowie[2]

[1] Department of Midwifery and Allied Health, Staffordshire University, UK, [2] NHS Education for Scotland, Glasgow, UK; Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK; School of Health, Science and Wellbeing, Staffordshire University, UK.

ABSTRACT

Situation Awareness (SA) is commonly defined 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, 1995:36). From this cognitive perspective, SA is synonymous with perception or attention, and involves a continual monitoring of the status quo for changes that might require action by frontline operators (Flin et al 2008). Within the midwifery literature, "loss of situation awareness" has been cited as a contributory factor to adverse events and unwanted clinical outcomes (HSIB 2020; Knight et al 2014; RCOG 2017). This operationalisation of SA is problematic for multiple reasons which are explored in this discussion paper.

The paper begins by exploring the transferability of human factors lessons between safety critical industries such as aviation and healthcare. Different theoretical perspectives on SA are evaluated, highlighting that the theoretical concept has been misapplied in midwifery, with distinct differences from Endsley's original model in how it is defined and measured. The paper provides an overview of the difficulties in measuring SA, which limit the prospective utility of the construct. Furthermore, retrospective identification of loss of SA is value laden and subject to hindsight bias. This stands in opposition to the Human Factors systems approach where "human error" should be viewed as a symptom of systemic problems within an organisation, rather than a causal factor (Amer-Wahlin and Dekker, 2008; Shorrock and Williams 2016).

This paper proposes that a more holistic perspective is required which considers the individual clinician within the context of the wider sociotechnical system, rather than focus solely on the performance of individuals. It is vital to identify the system factors which may lead to loss of situation awareness, in order to redesign the work environment to minimise patient harm and maximise safety (Singh et al 2006). Opportunity also exists for further research to investigate whether an alternative model of SA may be more appropriate for use in the healthcare context generally, and maternity care specifically, better reflecting the complex system in which clinicians work.

KEYWORDS

Situation Awareness, midwifery practice, maternity safety, Human Factors.

Background

Labour and birth are a time of great physiological change, with the potential for rapid deterioration in health of the mother and baby. During this time, the role of the midwife is to monitor maternal and fetal wellbeing to detect deviations from the expected course and act promptly to access emergency care where necessary (International Confederation of Midwives (ICM) 2017). However, Safety in UK maternity services is currently a high profile concern, with the Ockenden (2022) and the Kirkup (2015) reports identifying significant failings in care at two NHS Trusts, and several other Trusts under investigation or rated inadequate by the Care Quality Commission (CQC)(2022). Concerningly, it has been reported that improvements in care may have changed the outcome in 37% of cases of maternal death (Knight, 2021). Within the midwifery literature, "loss of situation awareness" has been cited as one contributory factor to adverse events and unwanted clinical outcomes in maternity care (Draper, Kurinczuk and Kenyon 2017:47; HSIB 2020; Knight et al 2014; RCOG

2017). However, the operationalisation of SA is problematic for multiple reasons, which will be explored below.

Learning lessons from other industries

The infamous American Institute of Medicine (1999) report To Err is Human: Building a Safer Health System, report cited alarming figures for avoidable patient deaths, referring to this as an “epidemic of medical errors” (Institute of Medicine 1999a:1). Their revolutionary conclusion was that medical errors are not due to poor practice by individual “bad apples” but result from system failures and conditions that cause people to fail (Institute of Medicine 1999b:49). Consequently, they recommended a move away from a punitive system of attributing blame, to redesigning work systems to support practitioners to do the right thing and recommended that human factors lessons be learnt from other safety critical industries to minimise errors and thus improve patient safety (Institute of Medicine 1999b).

Situation Awareness (SA) is an example of a human factors issue that has been highlighted in safety critical industries such as aviation, military, nuclear, and the oil and gas sectors, which has now been widely accepted in healthcare (Flin et al 2008; Gluyas and Harris 2016). Sharing learning between industries would seem logical in the pursuit of improving safety, however, concepts should not be indiscriminately transferred from one context to another (Powell-Dunford et al 2017). Concerning the application of situational awareness to healthcare generally, and midwifery specifically, it is important to consider the similarity or dissonance with aviation from whence the concept appears to have derived.

In a comparative review of aviation and healthcare, Kapur et al (2015) describe multiple differences between the industries. Some are obvious, such as that the focus of the work in healthcare is human bodies as opposed to inanimate aircraft, the relative lack of automation in healthcare and the fact that pilots will usually fly a specific type of airplane, whereas health professionals use a large range of different pieces of equipment to care for patients with a wide variety of different clinical presentations. Other less obvious differences are particularly relevant to situation awareness such as the small number of consistent crew members on an airplane, rather than large numbers of health professionals within a frequently changing team due to shift changes and work areas. Staff change-overs present an opportunity for information to be lost or may alter the team dynamics which could affect SA. Dekker (2011) uses a jet plane as an example of a complicated system, which are predictable, controllable, and stable when the correct procedure is followed. Complex systems such as midwifery on the other hand, are never fully “knowable”, they cannot be definitively mapped or measured because of the number of variables and dynamic interactions between elements (Dekker 2011:214). Complex systems are ones in which there are multiple interrelated components or agents, and the interaction between the components and the environment are continually in flux (Dekker 2011). Not only are there multiple practitioners involved in maternity care, but also, as Fioratu et al (2010) explain, the situation itself is not static, it is affected by the actions taken in response to the situation. Thus, the individual clinician is an integral part of the system, both responding to changes in the environment and also causing further change in a continuous cycle (Stanton et al 2010). This presents a clear obstacle to the study of SA in midwifery, where so many unknowable variables exist which may affect both individual clinicians and the system functioning as a whole. Furthermore, Kapur et al (2015) highlight hierarchical boundaries and cultural differences in healthcare, where aviation has a more embedded safety culture, free from blame. Acknowledging these differences is not to say that human factors principles, such as SA should not be applied to midwifery. Indeed, Fore and Sculli (2013:2619) argue that “SA needs to be examined in a theoretical context, studied systematically and openly recognised as a universal factor in patient safety”. However, the theoretical and practical relevance of SA to midwifery should be ascertained, and implementation ought to be tailored to the context (Powell-Dunford et al 2017).

Situation Awareness as a theoretical concept

Situation awareness is widely cited 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 1995:36). Within this cognitive perspective, SA is synonymous with perception or attention, it involves a continual monitoring of the status quo for changes that might require action (Flin et al 2008). This requires an "internalized mental model of the current state of the operator's environment" (Jones et al 2011:227), meaning that the practitioner integrates all of the available information into a picture of the current situation upon which decisions can be made. In this way, SA is contained within the mind of the individual practitioner. A key criticism of this perspective is that situating SA within the cognitive processing of the individual cannot account for behaviour of the system beyond that individual (Stanton et al 2010). As Fioratu et al (2010) explains, it is necessary to examine the interaction of the individual within the context of their environment, analysis should not focus solely on the individual.

Salmon et al (2009a) present an alternative social view of SA whereby Distributed Situation Awareness is the collective awareness of a whole system, with the cognition being shared across the system, within the interactions between system elements such as people and technology. Both Endsley's (1995) and Salmon's (2009a) models pertain to an awareness of the current state, where they diverge is upon where that awareness is held; in the mind of the individual practitioner, or somewhere between the elements of a socio-technical system. The DSA perspective aligns more closely with the human factors principle of systems thinking, whereby SA is distributed throughout the system, and viewed as a collaborative activity, rather than a largely individualistic one. Carayon et al (2014) highlight the importance of taking a systems approach to consider all aspects of the work system which may impact upon patient safety. Although Endsley's (1995) model does allude to external factors which may affect SA, such as system capability, interface design, complexity, and automation, these are extrinsic to situation awareness, rather than integral to it as in the DSA model by Salmon et al (2009a).

A third perspective exists, which situates SA purely within the devices that humans use as sources of information (Stanton et al 2010). It is questionable as to whether this perspective is actually describing SA or merely information availability. Indeed Stanton et al (2010) argue that more information does not necessarily lead to more situation awareness. Furthermore, this engineering perspective may hold weight in a purely technological system, however, is less applicable to healthcare and specifically to midwifery which as a socio-technical system, involves a large degree of human interaction, therefore SA cannot be contained only within devices in this context.

Measurement of Situation Awareness

It is important to note that irrespective of the debate between conceptual models, SA itself is a contentious issue. Within the human factors community, some authors have questioned whether SA exists as an entity (Salmon et al 2009b). Significant debate exists as to whether SA is an operational or representational concept (Endsley 2015; Dekker 2015) and indeed, viewing SA as a cognitive activity, as in Endsley's model, is fraught with issues when it comes to proving it exists in the world, beyond being a theoretical construct. Firstly, it is not possible to directly observe SA, therefore proxies must be used which are thought to demonstrate externally the processes that are going on in the mind of the individual (Flin et al 2008). For example, Abbott et al (2012) reported observable features of teamwork such as cooperation, co-ordination, leadership, monitoring and communication, and narratively equated these with levels of Team SA. Consequently, SA may be a subjective judgement imparted by the observer. Endsley (1988) has developed a measurement technique that is identified as an objective measure. The Situation Awareness Global Assessment Technique (SAGAT) uses a freeze-probe technique designed for use in simulated settings where the activity can be paused, and questions asked of participants (Endsley 1988). The generalisability of SAGAT findings to real world settings is limited given the complexity of clinical care environments with multiple other patients and the inability to freeze activity whilst clinicians consider the facets of the situation evolving in front of them. Additionally, the questions that are asked during the freezes of the SAGAT test knowledge rather than the cognitive processes used to obtain that knowledge (Chatzimichailidou et al 2015). Therefore, it could be argued that it is recall that is being tested moreover situation awareness. The Situation Present Assessment Technique (SPAM) technique is an alternative to SAGAT, which has been designed to be used in real time, without the need to freeze situations

(Durso et al 1998). However, this may not be practicable in midwifery practice because the SA probe questions could be disruptive to patient care or bias the results if clinicians choose to delay answering the probes until workload permits (Endsley 2021). Alternatively, Zhang et al (2020) provide a useful review of various physiological measures of situational awareness, such as eye tracking, cardiovascular changes and brain activity. Regardless of questions over the validity of these techniques given that physiological measures could be affected by environmental factors, the practicality of implementing the physiological measurement technologies whilst providing clinical care limits their use (Zhang et al 2020).

The implication of these difficulties in measuring SA are that we cannot be sure that the mental processes occur as depicted in the theoretical model, and if the depiction is accurate, then what distinguishes these activities as Situation Awareness rather than features of another theoretical construct such as attention or decision making for example. Endsley (2000) vehemently denies that decision making is synonymous with SA, SA is considered a precursor to it. Nonetheless, many of the barriers to SA identified in the literature pertain to cognitive limitations such as limitations of working memory capacity, attentional tunnelling, information or task overload, stress, fatigue, distraction and mind wander (Flin et al 2008; Gluyas and Harris 2016; Endsley 2015). Distributed Situation Awareness is more observable, as it is concerned with the interactions between elements of the system, however this is also subject to the same critique, that these interactions could be labelled communication or information sharing for example. It is debatable whether these interactions necessarily provide awareness to the clinician or simply information which must then be processed and acted upon. Fore and Sculli (2013) conclude that SA may be an amalgamation of similar terms. Therefore, the label given to the process could be considered semantic and may have little bearing on the day-to-day practice of clinicians. What is important is how the concept is applied in practice.

Use of Situational Awareness in Midwifery

A scoping review was undertaken to understand how the concept of SA is understood and used within the midwifery context. The review found that within midwifery, situation awareness is universally viewed as a person-level cognitive construct, dominated by Endsley's (1988) three step model. However, the theoretical concept has been misapplied in midwifery, with distinct differences from Endsley's original model in how it is defined and measured. For example, Rayfield et al (2017) define SA as a cognitive process, where Edozien (2015) concurs with Endsley (1995) that SA is a cognitive state. Endsley (1995: 36) refers to the process of acquiring SA as "situation assessment". This important because it determines how SA should be measured; by process or outcome measures, if indeed cognitive functions can be measured (Salmon et al, 2009a). Only two out of the six primary research studies included in the scoping review attempted to quantitatively measure SA. Both studies cited the SAGAT devised by (Endsley 1988), however the implementation of their methods deviated (Cooper et al, 2012; Morgan et al 2015). Two studies used qualitative methods which justifiably did not measure SA, yet made inferences about the levels of SA observed (Abbott et al 2012; Mackintosh et al 2009). Meanwhile, SA was a finding, rather than a prospective measure of the qualitative document analysis of delivery suite co-ordinator job descriptions by Bunford and Hamilton (2019). Interestingly, situation awareness did not feature in the job descriptions verbatim, this was an outcome of the logic modelling which was employed by the researchers. The final study measured participants' knowledge about the topic of SA pre- and post- training, without measuring SA itself (Sonesh et al 2015). Irrespective of the debate around the validity of measuring knowledge of the construct rather than participants' ability to demonstrate SA, the study found that knowledge was not improved by the training programme anyway (Sonesh et al 2015).

Prospective utility of Situation Awareness

Training has historically taken the form of either theoretical teaching about SA, or simulation-based practice using a crisis/ crew resource management approach of managing a clinical situation with factors that threaten SA added into the scenarios to challenge participants (Gordon et al 2012). However, given the difficulties in measuring SA, it is difficult to assess whether training is effective at improving individuals' SA. Furthermore, SA training does not appear from the literature to be linked to improved clinical outcomes (Fore and Sculli 2013). Endsley (2000) explains that the link between SA and performance is difficult because Clinicians can still make poor decisions even with good situation awareness. This rather questions the relevance of SA and the justification for attempting to teach it.

SA training is not appropriate when SA is viewed as a socio-technical construct, as the focus is the system and not the individual. The goal of Distributed Situational Awareness (DSA) is to improve the design of systems and technology to enable better human performance (Salmon et al 2009a). However, this is also challenging given that the DSA model describes SA but does not measure it (Chatzimichailidou et al (2015). Consequently, it is difficult to quantify any improvement in system performance in terms of SA, and impossible to attribute causation of any observed improvements to SA interventions. In this sense, the inability of the DSA model to measure SA renders itself practically impotent. This may be symptomatic of what Shorrock and Williams (2016: 97) call *"the inherent contradiction between human factors' pragmatic orientation and the systems approach it tries to build upon"*, in that a systems approach tries to look at problems holistically, but this means that the results are inevitably descriptive and less causal than those produced by a traditional reductionist approach which studies variables in isolation, in controlled environments. However, if the issue cannot be reduced to a single identifiable cause, then it is impossible to "fix the system" when elements of the system are interconnected and interdependent.

Retrospective assessment of Situational Awareness

The discussion thus far has demonstrated that SA is difficult to measure, difficult to teach with any real-terms effect, and too descriptive to be of use in quality improvement activities. Thus, there is limited prospective utility of SA theory. Within the midwifery literature, the concept of SA has been applied retrospectively, by citing loss of situation awareness as a contributory factor to adverse events and poor clinical outcomes (HSIB 2020; Knight et al 2014; RCOG 2017). This operationalisation of SA is problematic for a number of reasons.

The first issue is validity. If SA is a cognitive activity, then we cannot know what was occurring in the mind of the individual that led to the outcome under investigation. As outlined previously, measurement of SA is challenging in the present, therefore even more so in retrospect. In the case of retrospective identification of loss of SA, the assessment is made on the basis of observed decisions and actions, however this does not illuminate the cognitive process that was undertaken by the clinician to decide upon the ensuing actions. This raises another issue which is that SA is measured against a normative ideal which assumes that there was one correct understanding of the situation and therefore one course of action (Stanton et al 2017). However, there is not necessarily a "correct" course of action in Midwifery, there may be multiple routes to arrive at the same understanding or diagnosis (Singh et al 2006). Therefore, it is subject to expert opinion as to whether the clinician made appropriate decisions in any particular case. This is not necessarily a valid measure of the clinicians SA, moreover an appraisal of their decision making.

Furthermore, judging SA retrospectively is clearly subject to hindsight bias. With the benefit of hindsight, the eventual outcome is predictable, however at the time clinicians may not have had all of the information, or there may have been misleading symptoms that allude to an alternative pathology for example. Dekker (2015:159) suggests, *"'loss of situation awareness' is analytically nothing more than a post hoc judgment that says we know more about the situation now than other people apparently did back then"*. The human factors principle of local rationality says that people make decisions that make sense to them at the time. Although in retrospect an expert can judge that a decision was wrong, it must be assessed, based on the knowledge the practitioner had at the time and in the context that the decision was made. Furthermore, the human factors principle of performance variability suggests that people create safety by varying their practice to defend against potential threats within the system (Amer-Wahlin and Dekker 2008). What might superficially appear to have been a poor decision, may have been made with good reason, to compensate for inadequacies elsewhere within the system. Therefore, it is important to look at the systemic factors which ordinarily keep the organisation running safely, to understand why an adverse outcome may have occurred on this occasion (Dekker 2011).

The final concern with assessing SA retrospectively, it that there is an embedded value judgment. To assert that someone lacked SA, inherently implies a failure on the part of the clinician; that they got it wrong, that there was some form of negligence (Dekker 2015). This perspective stands in opposition to the human factors approach of systems thinking where human error should be viewed as a symptom of systemic problems within an organisation, rather than a cause (Amer-Wahlin and Dekker 2008). Blaming individuals, whether overtly or implicitly, is not helpful in bringing about solutions. Indeed, Shorrock and Williams (2016) argue that labelling errors retrospectively with broad terms such as loss of SA, removes context and so hinders our understanding of the system factors that were at play at the time. Therefore, identifying the reasons why someone may have lost SA could enable system redesign to reduce the risk of errors in the future (Singh et al 2006).

Conclusion

This paper has presented a theoretical critique of situation awareness as an academic concept, outlining the main perspectives that exist to explain this construct. A scoping review has demonstrated that despite heavy criticism of Endsley's (1988) cognitive model of SA, it appears that this perspective has been unquestionably implemented in midwifery settings, without consideration of alternative perspectives. Additionally, there are inconsistencies in the midwifery application of SA theory from Endsley's (1988) conceptual model, and how SA can be measured. Alarmingly, whilst assertions were made about levels of SA in the midwifery literature, only two studies attempted to measure it. Difficulties in measuring SA are a stumbling block for meaningful operationalisation of this concept in practice. If SA cannot be measured, then it cannot be taught because there it is impossible to evaluate whether the teaching has been effective. Furthermore, teaching SA does not appear from the literature to be linked to improved clinical outcome, therefore the benefit of teaching SA is debatable. In conclusion, SA may be an interesting and potentially useful theoretical concept, but the practical utility of it is limited. What is clear, is that applying "loss of situation awareness" post hoc as a cause or contributory factor in incident investigations and safety reports such as HSIB (2020), Draper et al (2017) and RCOG (2017), is not helpful in bringing about solutions (Shorrock and Williams 2016). Amer-Wahlin and Dekker (2008:936-7) argue *that "the greatest risk to safety in the delivery room is not the technology, nor the human. It is the oversimplification: the idea that there are simple explanations for adverse events and single silver bullets that can resolve the situation is an illusion"*. In line with the Institute of Medicine (1999b) recommendations set out at the beginning of this paper, which called for a move away from a punitive system of attributing blame, to redesigning work systems to support practitioners to do the right thing, it is vital to identify the system factors which may lead to loss of situation awareness, in order to redesign the work environment to minimise patient harm and maximise safety (Singh et al 2006).

References

Abbott, S., Rogers, M. and Freeth, D. (2012) 'Underpinning safety: Communication habits and situation awareness', British Journal of Midwifery, 20(4), pp. 279-284. doi: 10.12968/bjom.2012.20.4.279.

Amer-Wahlin, I. and Dekker, S. (2008) 'Fetal monitoring—a risky business for the unborn and for clinicians'. BJOG : an international journal of obstetrics and gynaecology, 115(8), pp. 935–937. doi:10.1111/j.1471-0528.2008.01758.x.

Bunford, D. and Hamilton, S. (2019) 'How delivery suite co-ordinators create situational awareness in the multidisciplinary team', British Journal of Midwifery, 27(8), pp. 497-505. doi: 10.12968/bjom.2019.27.8.497.

Carayon, P., Wetterneck, T.B., Rivera-Rodriguez, A.J., Schoofs Hundt, A., Hoonakker, P., Holden, R. and Gurses, A. P. (2014) Human factors systems approach to healthcare quality and patient safety. Applied Ergonomics, 45, pp 14-25.

Care Quality Commission (CQC) (2022) Find and compare services. https://www.cqc.org.uk/search/all?query=maternity&location-query=&radius=10&display=list&sort=relevance&last-published=&filters[]=archived:active&filters[]=lastPublished:all&filters[]=more_services:all&filters

[]=overallRating:Inadequate&filters[]=services:hospital&filters[]=specialisms:all (Accessed 14/07/2022).
Chatzimichailidou, M.M., Protopapas, A. and Dokas, I.M. (2015) Seven Issues on Distributed Situation Awareness Measurement in Complex Socio-technical Systems. In: Boulanger F., Krob D., Morel G., Roussel JC. (eds) Complex Systems Design & Management. Springer, Cham. https://doi.org/10.1007/978-3-319-11617-4_8 (Accessed 03/02/2022).
Cooper, S., Bulle, B., Biro, M.A., Jones, J., Miles, M., Gilmour, C., Buykx, P., Boland, R., Kinsman, L., Scholes, J. and Endacott, R. (2012) 'Managing women with acute physiological deterioration: Student midwives performance in a simulated setting', Women and Birth, 25(3), pp. e27-e36. doi: 10.1016/j.wombi.2011.08.009.
Dekker, S. (2011) Patient Safety: A Human Factors Approach. Taylor & Francis Group, ProQuest Ebook Central. http://ebookcentral.proquest.com/lib/staffordshire/detail.action?docID=726851 (Accessed 02/02/2022).
Dekker, S.W.A. (2015) The danger of losing situation awareness. Cogn Tech Work, 17: pp159–161.
Draper, E.S., Kurinczuk, J.L. and Kenyon, S. (2017) MBRRACE-UK Perinatal Confidential Enquiry. Term, singleton, intrapartum stillbirth and intrapartum-related neonatal death. https://www.hqip.org.uk/wp-content/uploads/2018/02/mbrrace-uk-perinatal-confidential-enquiry-report-2017.pdf (Accessed 28/02/2022).
Durso, F.T., Hackworth, C.A. and Truitt, T. et al (1998) Situation awareness as a predictor of performance in en route air traffic controllers. Air Traffic Quarterly, 6(1): 1-20.
Edozien, L. (2015) Situational Awareness and Its Application in the Delivery Suite. Obstetrics & Gynecology, 125, (1): 65-69.
Endsley, M.R. (1988) Situation awareness global assessment technique (SAGAT). Proceedings of the IEEE 1988 National Aerospace and Electronics Conference, pp. 789-795 vol.3, doi: 10.1109/NAECON.1988.195097. https://ieeexplore.ieee.org/document/195097 (Accessed 12/01/2022).
Endsley, M.R. (1995) Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors Journal, 37(1), 32-64.
Endsley, M. (2015) Situation awareness: operationally necessary and Scientifically grounded. Cogn Tech Work, 17:163–167. DOI 10.1007/s10111-015-0323-5.
Endsley, M.R. (2000) Theoretical Underpinnings of Situation Awareness: A Critical Review. In: Endsley, M. R. and Garland D. J (Eds.) (2000) Situation Awareness Analysis and Measurement. Mahwah, NJ: Lawrence Erlbaum Associates.
Endsley, M.R. (2021) Systematic Review and Meta-Analysis of Direct Objective Measures of Situation Awareness: A Comparison of SAGAT and SPAM. Human Factors, 63(1) 124-150.
Fioratu, E., Flin, R., Glavin, R. and Patey, R. (2010) Beyond monitoring: distributed situation awareness in anaesthesia. British Journal of Anaesthesia, 105 (1): 83–90 (2010). doi:10.1093/bja/aeq137.
Flin, R., O'Connor, P. and Crichton, M. (2017) Safety at the Sharp End A Guide to Non-Technical Skills. 1st edition. London: CRC Press.
Fore, A.M. and Sculli, G.L. (2013) A concept analysis of situational awareness in nursing. Journal of Advanced Nursing, 69(12), 2613–2621. doi: 10.1111/jan.12130
Gluyas, H. and Harris, S-J. (2016) Understanding situation awareness and its importance in patient safety. Nursing Standard, 30(34): 50-58.
Gordon, M., Darbyshire, D. and Baker, P. (2012) Non-technical skills training to enhance patient safety: a systematic review. Medical Education, 46: pp1042–1054.
Healthcare Safety Investigation Branch (HSIB) (2020) Delays to intrapartum intervention once fetal compromise is suspected. Independent report by the Healthcare Safety Investigation Branch I2019/020. https://www.hsib.org.uk/investigations-and-reports/delays-to-intrapartum-intervention-once-fetal-compromise-is-suspected/ (Accessed 17/01/2022).
Institute of Medicine (1999a) To Err is Human: Building a Safer Health System. Brief Report. https://www.nap.edu/resource/9728/To-Err-is-Human-1999--report-brief.pdf (Accessed 02/02/2022).

Institute of Medicine (1999b) To Err is Human: Building a Safer Health System. Full book. https://www.nap.edu/download/9728 (Accessed 03/02/2022).
International Confederation of Midwives (2017) Core Document: International Definition of the Midwife. https://www.internationalmidwives.org/assets/files/definitions-files/2018/06/eng-definition_of_the_midwife-2017.pdf (Accessed 02/03/2022).
Jones, R.E.T., Connors, E.S. and Endsley, M.R. (2011) A Framework for Representing Agent and Human Situation Awareness. 2011 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), Miami Beach, FL. https://ieeexplore-ieee-org.ezproxy.staffs.ac.uk/stamp/stamp.jsp?tp=&arnumber=5753450&tag=1 (Accessed 03/02/2022).
Kapur, N., Parand, A., Soukup, T., Reader, T. and Sevdalis, N. (2016) Aviation and healthcare: a comparative review with implications for patient safety. Journal of the Royal Society of Medicine Open. DOI:10.1177/2054270415616548. (Accessed 03/02/2022).
Kirkup, B. (2015) The Report of the Morecambe Bay Investigation. An independent investigation into the management, delivery and outcomes of care provided by the maternity and neonatal services at the University Hospitals of Morecambe Bay NHS Foundation Trust from January 2004 to June 2013. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/408480/47487_MBI_Accessible_v0.1.pdf (Accessed 24/01/2022).
Knight, M., Kenyon, S., Brocklehurst, P., Shakespeare, J. and Kurinczuk, J.J. (2014) Saving Lives, Improving Mothers' Care. Lessons learned to inform future maternity care from the UK and Ireland Confidential Enquiries into Maternal Deaths and Morbidity 2009-2012. MBRRACE Report. https://www.npeu.ox.ac.uk/assets/downloads/mbrrace-uk/reports/Saving%20Lives%20Improving%20Mothers%20Care%20report%202014%20Full.pdf (Accessed 10/02/2022).Powell-Dunford, N., Brennan, P.A., Peerally, M.F., Kapur, N., Hynes, J. and Hodkinson, P.D. (2017) Mindful Application of Aviation Practices in Healthcare. Aerospace Medicine and Human Performance, 88 (12): pp1107-1116.
Knight, M., Bunch, K., Tuffnell, D., Patel, R., Shakespeare, J., Kotnis, R., Kenyon, S. and Kurinczuk, J.J. (Eds.) (2021) MBRRACE-UK. Saving Lives, Improving Mothers' Care Lessons learned to inform Maternity care from the UK and Ireland Confidential Enquiries into Maternal Deaths and Morbidity 2017-19. https://www.npeu.ox.ac.uk/assets/downloads/mbrrace-uk/reports/maternal-report-2021/MBRRACE-UK_Maternal_Report_2021_-_FINAL_-_WEB_VERSION.pdf (Accessed 24/01/2022).
MacKintosh, N., Berridge, E.-. and Freeth, D. (2009) 'Supporting structures for team situation awareness and decision making: Insights from four delivery suites', Journal of evaluation in clinical practice, 15(1): 46-54. doi: 10.1111/j.1365-2753.2008.00953.x.
Morgan, P., Tregunno, D., Brydges, R., Pittini, R., Tarshis, J., Kurrek, M., DeSousa, S. and Ryzynski, A. (2015) 'Using a situational awareness global assessment technique for interprofessional obstetrical team training with high fidelity simulation', Journal of Interprofessional Care, 29(1):13-19. doi: 10.3109/13561820.2014.936371.
Ockenden, D. (2022) Findings, Conclusions and Essential Recommendations From the Independent Review of Maternity Services at the Shrewsbury and Telford Hospital NHS Trust. https://www.ockendenmaternityreview.org.uk/wp-content/uploads/2022/03/FINAL_INDEPENDENT_MATERNITY_REVIEW_OF_MATERNITY_SERVICES_REPORT.pdf (Accessed 14/07/2022).
Powell-Dunford, N., Brennan, P.A., Peerally, M.F., Kapur, N., Hynes, J. and Hodkinson, P.D. (2017) Mindful Application of Aviation Practices in Healthcare. Aerospace Medicine and Human Performance, 88 (12):1107-1116.
Rayfield et al (2017) Rayfield, M., Ansari, S. and Prosser-Snelling, E. (2017) Seeing the bigger picture. https://www.rcm.org.uk/news-views/rcm-opinion/seeing-the-bigger-picture/ (Accessed 06/01/2022).

Royal College of Obstetricians and Gynaecologists (RCOG) (2017) Each Baby Counts. 2015 Full Report. https://www.rcog.org.uk/en/guidelines-research-services/audit-quality-improvement/each-baby-counts/ebc-2015-report/ (Accessed 17/01/2022).

Salmon, P.M., Stanton, N.A., Walker, G.H., Jenkins, D., Ladva, D., Rafferty, L. and Young, M. (2009a) Measuring Situation Awareness in complex systems: Comparison of measures study. International Journal of Industrial Ergonomics, 39: 490–500.

Salmon, P.M., Stanton, N.A., Walker, G.H. and Jenkins, D.P. (2009b) Distributed Situation Awareness. Theory, Measurement and Application to Teamwork. London: CRC Press.

Shorrock, S. and Williams, C. (2016) Human Factors and Ergonomics in Practice: Improving System Performance and Human Well-Being in the Real World. Taylor & Francis Group, ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/staffordshire/detail.action?docID=5185433. Accessed 09/02/2022)

Singh, H., Petersen, L.A. and Thomas, E.J. (2006) Understanding diagnostic errors in medicine: a lesson from aviation. Qual Saf Health Care, 15:159–164. doi: 10.1136/qshc.2005.016444.

Sonesh, S.C., Gregory, M.E., Hughes, A.M., Feitosa, J., Benishek, L.E., Verhoeven, D., Patzer, B., Salazar, M., Gonzalez, L. and Salas, E. (2015) 'Team training in obstetrics: A multi-level evaluation', Families, Systems, & Health, 33(3): 250-261. doi: 10.1037/fsh0000148; 10.1037/fsh0000148.supp (Supplemental).

Stanton, N.A. , Salmon, P.M., Walker, G.H. and Jenkins, D.P. (2010) Is situation awareness all in the mind? Theoretical Issues in Ergonomics Science, 11:1-2: 29-40. DOI: 10.1080/14639220903009938.

Stanton, N.A., Salmon, P.M., Walker, G.H., Salas, E. and Hancocke, P.A. (2017) State-of-Science: Situation Awareness in individuals, teams and systems. Ergonomics, 60(4):247-258.

Zhang, T., Yang, J. and Liang, N. et al. (2020) ‘Physiological Measurements of Situation Awareness: A Systematic Review’, Human factors, pp. 1872082096907–18720820969071. DOI:10.1177/0018720820969071. Available at: https://doi.org/10.1177/0018720820969071.

Usability evaluation: an investigation on combination of analytical and empirical methods

Setia Hermawati, Glyn Lawson

Human Factors Research Group, University of Nottingham

SUMMARY

This study investigated the effect of combining analytical (heuristics and cognitive walkthrough) and empirical methods in usability evaluation. Data from two usability studies were used to simulate the outcomes of different combinations of usability evaluation methods. The findings show that the combined analytical methods significantly reduce the number of participants required in the empirical method without compromising the results of the usability evaluation.

KEYWORDS

Usability evaluation, Heuristics evaluation, Cognitive walkthrough, User testing

While empirical evaluation (user testing) is desired as part of usability evaluation, it is often costly and cumbersome to conduct (Nielsen, 1993) because they require the need to recruit test participants that match the target user for a system. Analytical methods, also called discounted methods, were developed to reduce the costs of usability evaluation associated with empirical methods and involve the participation of experts in human factors. Heuristic evaluation is cheap, fast and able to predict major usability problems that could potentially occur during usability testing (Jeffries et al., 1991; Tang et al., 2006; Hwang and Salvendy, 2010). However, it is also reported to often discover low-priority usability problems, and its output is largely dependent on the quality of the evaluators involved (Jeffries and Desurvire, 1992; Hwang and Salvendy, 2010). Another analytical method of interest is cognitive walkthrough. It is costlier than heuristic evaluation and requires extensive knowledge in cognitive psychology (Bias and Mayhew, 2005; Hwang and Salvendy, 2010). Although it is more effective in finding severe problems (Sears, 1997), it can reveal only about a third of the usability problems detected by a heuristic evaluation (Jeffries et al., 1991).

Several studies have compared the performance of analytical and empirical methods (see, e.g., Karat et al., 1992; Ahmed, 2005; Tan et al., 2009; Thyvalikakath et al., 2009; Petri and Power, 2012). Although general findings suggest that analytical methods can identify usability problems that severely affect interaction with a system, most studies found that it is very unlikely that analytical methods alone can identify all severe usability problems. There is an argument that the analytical methods should not be used to justify the omission of the empirical methods as part of usability evaluation (Jeffries and Desurvire (1992) and that a combination of empirical methods and analytical methods should be adopted (Ahmed, 2005); Tan et al., 2009).

Unfortunately, there has not been a study exploring how combinations of analytical and empirical methods affect the discovery rate of usability problems. This study was aimed to fill this gap and sought to provide evidence of the benefit of combining usability evaluation methods. To achieve this aim, we conducted a simulation study that was based on data from actual usability studies to fully investigate the interaction among the different usability evaluation methods with respect to the number of unique usability issues that could be identified and the risk of missing severe usability issues.

We conducted two independent usability evaluation studies on two different software. The software was designed for educational/training purposes with the first software aimed at trainers of assembly line operators in a manufacturing setting, and the other one was aimed at students in a higher education setting. Two different and yet similar themed software were intentionally used as it would allow, to some extent, generalisation of the outcome of this study. In each study, three types of usability evaluation methods (heuristic, cognitive, and empirical) were conducted, and five participants were assigned in each evaluation method. The heuristics and cognitive walkthrough involved participants who completed the graduate or post-graduate level of coursework's in Human Factors. In the empirical evaluation, participants consisted of trainers of assembly line operators and students at the Nottingham University.

For each study, the severity of each usability issue was identified. Next, the usability issues were grouped and coded to remove redundancies. The coded usability issues were then assigned to each participant. After this step was completed, a simulation of hypothetical groups that represented variations of all participants and number of participants in each method was performed. For each study, a total of 150 groups were created, combining different numbers of participants in each evaluation method. In each group, subgroups were then created to reflect different combinations of participants. The creation of subgroups ensured that the simulation considered differences in the performance of the participants in identifying the usability problem. Between 5 to 1000 subgroups were created in each group. The simulation was achieved by creating a programming code in MATLAB.

The findings of this study showed that, in comparison to heuristic method, cognitive walkthrough identified more unique usability problems and resulted in a better prediction of usability problems that would be encountered by end-users in empirical method. Furthermore, we also found that the combination of cognitive and heuristic methods identified 98.2% of known usability problems, compared to 44.8% and 68.9% by heuristic and cognitive methods, respectively. Although this finding is promising and suggests the potential of combined analytical methods, this study also found that there was still a risk of missing severe usability issues when usability evaluation relied solely on analytical evaluation methods. Regarding the question 'how many participants are required in the empirical method to compensate for combined analytical methods', our study showed that the participation of just one participant in the empirical method complemented combined analytical methods and successfully reduced the risk of missing severe usability problems to less than 75%.

The results of this study also revealed that, at least nine participants (4 and 5 usability experts to conduct heuristics and cognitive walkthrough, respectively) were required in combined analytical methods to identify 85% of unique usability problems. This finding disputed assertions that about 85% of known usability problems could be identified by 5 participants. This study also found that the diminishing return relationship between the number of participants in empirical method and the number of usability issues identified was also applicable to combined analytical methods. This suggests that Nielsen and Landauer's (1993) equation that illustrates the relationship between the discovery of usability problems and the number of participants could also be used to estimate the total number of participants required in combined analytical methods.

References

Ahmed, S. M. Z. (2008) A comparison of usability techniques for evaluating information retrieval system interfaces. Performance Measurement and Metrics. 9(1), 48-58.

Bias, R. G., & Mayhew, D. J. (2005) Cost-justifying usability: an update for the internet age. San Francisco: Morgan Kaufman.
Hwang, W., & Salvendy, G. (2010) Number of people required for usability evaluation: the 10±2 rule. Communications of the ACM, 53(5), 130-133.
Jeffries, R., & Desurvire, H. (1992) Usability testing vs. heuristic evaluation: was there a contest. ACM SIGCHI Bulletin, 24(4), 39-41.
Jeffries, R., Miller, J.R., Wharton, C. & Uyeda, K. (1991) User interface evaluation in the real world: a comparison of four techniques. In Proceedings of the SIGCHI conference on Human Factors in computing systems: Reaching through technology (pp. 119-124).
Karat, C-M., Campbell, R., & Fiegel, T. (1992) Comparison of empiric testing and walkthrough methods in user interface evaluation. In Proceedings of CHI 1992 (pp. 397-404).
Nielsen, J. (1993) Usability Engineering. Cambridge, MA: Academic Press.
Petri, H. & Power, C. (2012) What do users really care about? A comparison of usability problems found by users and expert on highly interactive websites. In Proceedings of CHI 2012 (pp. 2107-2116).
Sears, A. (1997) Heuristic walkthroughs: finding the problems without the noise. International Journal of Human-Computer Studies Interaction, 9(3), 213–234.
Tan, W-s., Liu, D., & Bishu, R. (2009) Web evaluation: heuristic evaluation vs. user testing. International Journal of Industrial Ergonomics, 39, 621-627.
Tang, Z., Johnson, T.R., Tindall, R.D., & Zhang, J. (2006) Applying heuristic evaluation to improve the usability of a telemedicine system. Telemed J E Health, 12(1), 24–34.
Thyvalikakath, T. P., Monaco, V. M., Thambuganipalle, H., & Schleyer, T. (2009) Comparative study of heuristic evaluation and usability testing methods. Studies in Health Technology and Informatics, 143, 322-327.

Driving at night and how it’s influenced by perceived driver skills

İbrahim Öztürk[1] & Natasha Merat[1]

[1]Institute for Transport Studies, University of Leeds, UK

SUMMARY

Night-time driving is associated with higher crash rates, partly due to reduced visibility of the driving environment. Crash data show that young drivers are over-represented in night-time incidents, world-wide. Self-reported driver skills provide an important indicator of driving abilities, which can also interact with night-time driving performance. This study investigated the relationship between drivers’ self-reported driver skills and their perceived night-time driving challenges, comparing responses between young and older drivers. Results found that young drivers with lower self-reported perceptual-motor skills and higher safety skills experienced more difficulties associated with night-time driving.

KEYWORDS

self-reported driver skills, night-time driving, young drivers, older drivers, driving difficulties

Introduction

Night-time driving can be challenging due to limited visibility of the driving environment and problems with glare from other vehicles (e.g., Evans et al., 2020; 2022). Research also shows that inexperienced young drivers are less likely to detect hazards compared to experienced older drivers (Borowsky et al., 2010), and are more likely to be involved in crashes at night in the UK (Regev et al., 2018). When compared to young and less experienced drivers, experienced and relatively older drivers are reported to be more aware of night-driving challenges, for example, those associated with reduced visibility (Evans et al., 2022).

Drivers’ performance is influenced by a combination of their behaviour and skills. While behaviour (style) is “*the way individuals choose to drive or driving habits that have become established over a period of years*”, driver skills (performance) “*limits to performance on elements of the driving task”* (Elander et al., 1993, p.279). In other words, the former one explains what the driver “does” and the latter concerns what the driver “can do” (Özkan & Lajunen, 2011). According to the “two pathways to a crash” model (Lajunen & Özkan, 2021), behaviours and skills influence crash involvement through driver violations and errors, respectively. More specifically, self-reported driver skills, as an indicator of drivers’ abilities and performance, are a crucial predictor of unsafe behaviours and outcomes, including crashes, and are shown to be affected by experience and general cognitive abilities (e.g., Lajunen & Özkan, 2021; Xu et al., 2018).

Whilst few researchers have investigated the risks experienced by young (e.g., Evans et al., 2020; Regev et al., 2018) and older (e.g., Kimlin et al., 2020; Wood, 2019) drivers at night, to the best of our knowledge, it is not known how night-time driving difficulties vary, based on age and perceived level of driver skills. Therefore, the present study collected self-reported driver skills and night-time driving difficulty data from two groups of drivers (young, older), to understand how these accounts are affected by participant age and gender. The following research questions were addressed in this study:

1) Are there any age and gender differences in self-reported driver skills and night-time driving difficulties?
2) How does the relationship between age and difficulties in night-time driving vary as a function of perceptual-motor skills by safety skills?

Method

Participants

Sixty participants were invited to take part in the study, and 57 participants completed the self-report sections. Three participants were excluded due to technical problems. The cohort included 30 young drivers (15 male, 15 female) between 21 and 25 years old (M = 22, SD = 1) and 27 older drives (18 male, 9 female), aged between 59 and 79 years (M = 66, SD = 4).

Measures

Driver Skills Inventory (DSI): Self-reported driver skills were measured with the 20-item Driver Skills Inventory (Lajunen & Summala, 1995). Self-reported driver skills were conceptualised under two dimensions: perceptual-motor skills and safety skills, and measured with ten items for each. While perceptual-motor skills focus on control aspects of driving, safety skills are related to drivers' safety motivation (Lajunen & Summala, 1995). Drivers were asked to indicate how weak/strong the 20 aspects of driving were on a 5-point Likert from definitely weak (0) to definitely strong (4). The Cronbach's alpha reliabilities of the subscales were .79 for perceptual-motor skills and .86 for safety skills.

Vision and Night Driving Questionnaire (VNDQ): Drivers' self-reported visual difficulties during night-time driving were measured with a 9-item scale, developed by Kimlin et al. (2016). Participants were asked to indicate how difficult it was to complete nine different tasks during night-time driving, using a 5-point Likert scale from no difficulty (0) to extreme difficulty (4). The scale corresponded to a single factor with .88 Cronbach's alpha reliability.

Procedure

The two groups of participants were recruited for a driving simulator study, approved by the ethics committee of the School of Business, Environment and Social Services, University of Leeds (AREA 21-108) as part of the EPSRC–funded HAROLD project (EP/S003576/1). Convenient and snowball sampling methods were used to recruit participants, and social media accounts were also used to advertise the study. Eligible participants (who had a valid UK driving license, were regular drivers, drove at least once a week, and had normal or corrected to normal vision) were invited to take part. The main aim of the study was to investigate how driver behaviour during day and night was affected by a cognitively loading non-visual n-back task (not reported here, see Öztürk et al., 2023). After completing the last drive, participants were asked to complete the two questionnaires reported above, which are reported in this paper.

Analysis

First, Analyses of Variance (ANOVA) were used for the Vision and Night Driving Questionnaire items, to investigate the differences between young and older drivers, for each of the different tasks. In the second step, any difference in age and sex for driver skills and night-time driving difficulties were explored using a 2 (Age: young, older) by 2 (Sex: female, male) Analysis of Covariance (ANCOVA) with 5000 bootstraps, where annual mileage was entered as a control variable. In the third step, the moderated moderation analysis by Hayes (2022) was conducted, to explore the relationship between age and difficulties in night-time driving, and whether these are influenced by drivers' self-reported perceptual-motor skills and safety skills (Figure 1). The analyses were performed by using the PROCESS macro model 3 for SPSS, with 5000 bootstraps, controlling for sex and annual mileage. Perceptual-motor skills and safety skills were entered into the model as mean-centred variables. The statistical significance value was determined as .10, considering the low statistical power for interaction effects (Morris et al., 1986). Significant interaction effects were shown by using three values of moderators as the mean and one standard deviation above and below the mean.

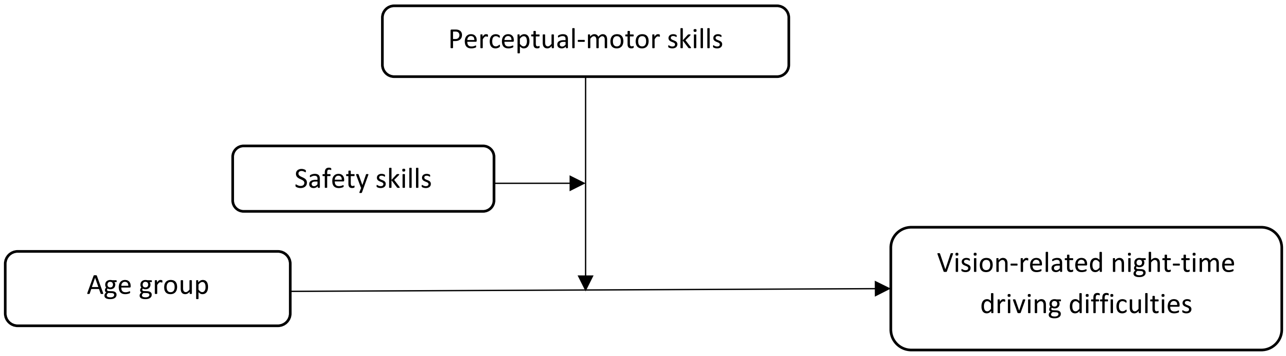


Figure 1. Conceptual diagram of moderated moderation analyses

Results

Night-time driving difficulties

In terms of Vision and Night Driving Questionnaire, when comparing the two groups, young drivers perceived "*Seeing pedestrians or animals on the road side*" as more difficult than older drivers. While the overall level of perceived difficulty for each task was not high, for young drivers the most difficult task was "*Seeing the road in rain or poor weather*", while "*Seeing because of glare when driving at dusk or dawn"* was reported as most difficult for older drivers. *"Judging the distance between you and other moving cars"* while driving at night was perceived to be the least difficult task for both groups of drivers (Table 1).

Table 1. Descriptive statistics for the different tasks rated for night-time driving

	Young		Older		F	*p*
	M	*SD*	*M*	*SD*	(1,55)	
Seeing pedestrians or animals on the roadside when driving at night	1.77	.86	1.19	.79	7.06	.010
Seeing the road in rain or poor weather when driving at night	2.00	1.05	1.52	.96	3.19	.079
Seeing the road because of oncoming headlights when driving at night	1.97	1.03	1.70	.78	1.16	.286
Adjusting after passing headlights from oncoming cars when driving at night	1.27	.98	1.41	.93	.31	.582
Seeing because of glare when driving at dusk or dawn	1.97	.93	1.85	.86	.23	.632
Judging the distance between you and other moving cars while driving at night	.90	.76	.81	.83	.16	.688
Judging the distance to your turnoff or exit while driving at night	.90	.71	.96	.71	.11	.739
Seeing dark coloured cars when driving at night	1.30	.53	1.26	.71	.06	.807
Reading street signs when driving at night	.97	.85	.93	.87	.03	.859

Differences in perceptual-motor and safety skills

For perceptual-motor skills (Table 2), the main effect of sex was significant ($F(1, 52) = 16.63$, $p < .001$, $\eta^2_p = .24$), with male drivers revealing higher self-reported perceptual-motor skills than female drivers. There was no significant difference in terms of age ($F(1, 52) = .06$, $p = .813$, $\eta^2_p = .00$), or the interaction of age and sex ($F(1, 52) = .72$, $p = .401$, $\eta^2_p = .01$).

For safety skills (Table 2), the main effect of age was significant ($F(1, 52) = 7.93$, $p = .007$, $\eta^2_p = .13$). Older drivers reported higher safety skills than young drivers. No significant sex difference ($F(1, 52) = 1.79$, $p = .187$, $\eta^2_p = .03$) or interaction effect ($F(1, 52) = .02$, $p = .900$, $\eta^2_p = .00$) was observed for safety skills.

For night-time driving difficulties (Table 2), no differences were observed for either age ($F(1, 52) = 1.36$, $p = .249$, $\eta^2_p = .03$), or sex ($F(1, 52) = .17$, $p = .685$, $\eta^2_p = .00$). The interaction of age and sex ($F(1, 52) = .92$, $p = .341$, $\eta^2_p = .02$) was also not significant.

Table 2. Descriptive statistics for perceptual-motor skills, safety skills and vision-related difficulties, by age and sex

		Perceptual-motor skills		**Safety skills**		**Night-time driving difficulties**	
		M	***SD***	***M***	***SD***	***M***	***SD***
Young	**Male**	3.10	.35	2.37	.49	1.39	.63
	Female	2.51	.46	2.54	.54	1.51	.45
	Total	2.80	.50	2.46	.52	1.45	.54
Older	**Male**	2.97	.45	2.78	.59	1.37	.58
	Female	2.58	.41	3.04	.61	1.10	.88
	Total	2.86	.47	2.85	.60	1.29	.67
Total	**Male**	3.03	.41	2.60	.58	1.38	.59
	Female	2.54	.44	2.71	.60	1.37	.64
	Total	2.83	.48	2.65	.59	1.37	.61

The effect of age and driver skills on night-time driving difficulties

The moderated moderation model was found to be significant ($F(9, 47) = 2.47$, $p = .021$), and explained 32% of the variance (Table 3). The three-way interaction between age, perceptual-motor skills, and safety skills contributed to 6% of the additional variance ($F(1, 47) = 4.30$, $p = .044$).

Table 3. Conditional effects of age group on vision-related night-time driving difficulties by perceptual-motor skills and safety skills

Variable	*b*	*se*	*t*	*p*	95% CI
Age group (1: Older, 2: Young)	.22	.17	1.30	.202	-.12, .55
Perceptual-motor skills	-.67	.63	-1.05	.297	-1.94, .61
Age by perceptual-motor skills	-.05	.37	-.13	.896	-.79, .69
Safety skills	-.54	.44	-1.22	.228	-1.43, .35
Age by safety skills	.29	.28	1.04	.305	-.27, .85
Perceptual-motor skills by safety skills	2.13	.97	2.19	.034	.17, 4.09*
Age by perceptual-motor skills by safety skills	-1.41	.68	-2.07	.044	-2.78, -.04*
Sex (1: Male, 2: Female)	-.29	.19	-1.56	.126	-.67, .08
Annual mileage	.00	.00	1.88	.067	.00, .00

Note. * Significant effect on night-time driving difficulties

There was a positive interaction between age at the low level of self-reported perceptual-motor skills and high level of safety skills ($b = .81$, $t(1, 47) = 2.30$, $p = .026$), with younger drivers reporting more difficulties in night-time driving, compared to older drivers (Figure 2).

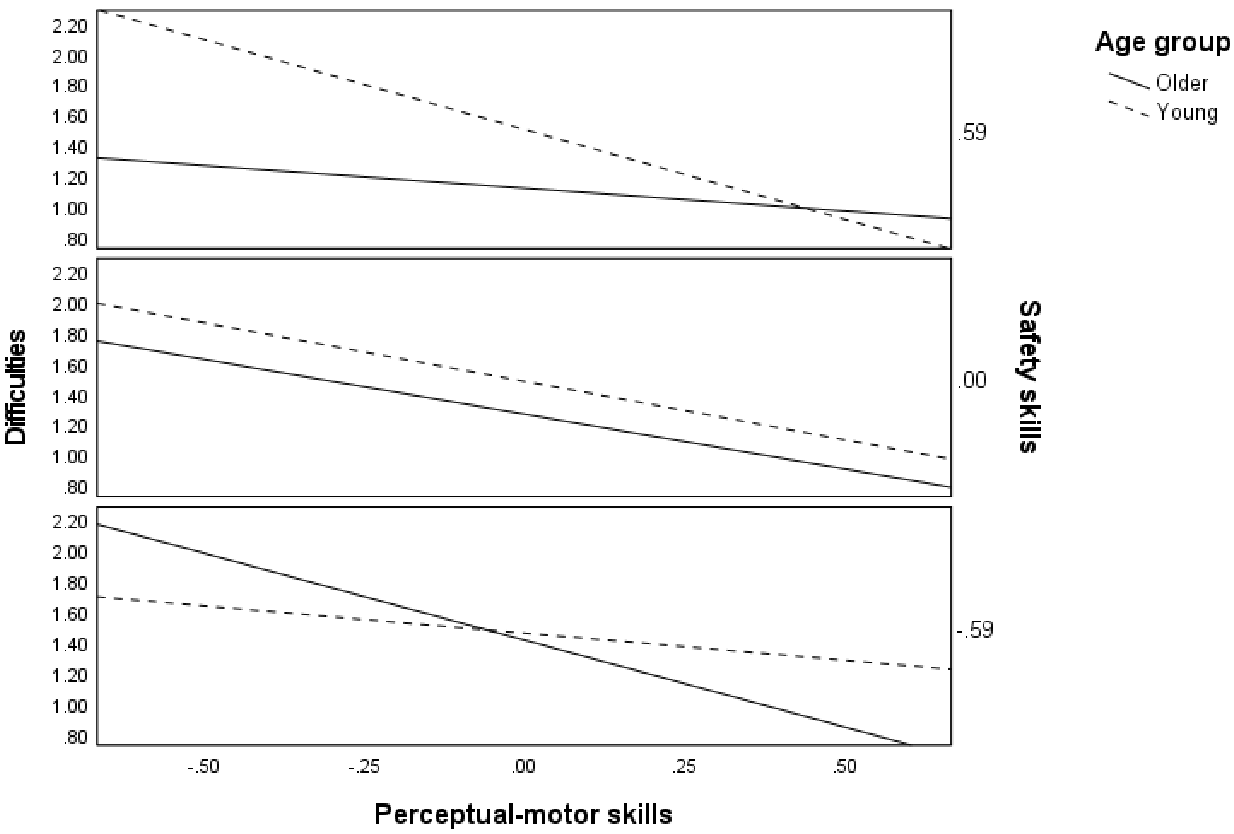


Figure 2: The relationship between age and night-time driving difficulties as a function of perceptual-motor skills by three levels (high, moderate, low) of safety skills

Discussion

The aim of this study was to investigate if there are any differences between young and older, and/or male and female drivers, in terms of their self-reported driving skills, and night-time driving difficulties. The study also sought to investigate whether the difficulties experienced when driving at night differed as a function of the perceived level of perceptual-motor and safety skills. Previous work has shown that drivers reporting higher perceptual-motor skills are also less likely to commit fewer errors and lapses (Xu et al., 2018), while high safety skills are negatively correlated with violations and penalty points (Xu et al., 2018). This work was used to as a basis for investigating the interactions between drivers' self-reported skills, to explain night-time difficulties experienced by young and older drivers.

Regarding the first aim of the study, older drivers declared more safety skills when compared to young drivers, which is in line with previous studies (Martinussen et al., 2014; Ostapczuk et al., 2017, Xu et al., 2018). Male drivers in our study also reported higher perceptual-motor skills than female drivers. However, contrary to previous studies (e.g., Gruber et al., 2013; Kimlin et al., 2020), the perceived level of difficulty across different driving tasks at night was low and was not different for our two groups of drivers. This low level of perceived difficulty in night-time driving may be due to high optimism bias (DeJoy, 1989; White et al., 2011) and an overestimation of the individuals' own driving abilities, or a lack of self-awareness of their driving skills (e.g., Freund et al., 2005; Martinussen et al., 2017; McKenna et al., 1991; Parker et al., 2001). On the other hand, in agreement with previous work (Evans et al., 2022), our study found that young drivers were less aware of their limited visibility issues during night-time driving, which can account for their overrepresentation in night-time crashes (Regev et al., 2018). Further work in this area is therefore warranted.

Regarding the second aim of this study, the three-way interaction effect showed that, when self-reported perceptual-motor skills were low and safety skills were high, younger drivers reported more difficulties at night-time than older drivers. For this group of young drivers, high safety motives may mean that they are more aware of night-time driving difficulties, which they associate with their with lower perceptual-motor and technical skills, likely related to their relatively lower driving experience (De Craen et al., 2011; Martinussen et al., 2014; Ostapczuk et al., 2017). Driver training that focusses on improving the night-time driving experience of this group may, therefore, be of value.

Regarding limitations of the study, although the sample size is sufficient to provide the proposed relationships, a bigger sample size would improve confidence in the results, and our conclusions regarding self-reported difficulties associated with night-time driving. Additionally, although anonymity and confidentiality were ensured in this study, results may have been biased by social desirability (Yilmaz et al., 2022), or participants' own evaluation of their driver skills, including an overestimation of their capabilities (McKenna et al., 1991).

These results have important implications for developing interventions to improve safety among both young and older drivers. These may include training programs focusing on the risks associated with night-time driving to overcome potential effects of optimism bias or overestimation of driving skills, and improving night-time driving skills, especially for young drivers, to reduce the difficulties experienced and the risk of near-misses and crashes. The potential benefits of driver skills may also be beneficial for older drivers, including refresher courses to keep their skills up-to-date (e.g., older drivers and Advanced Driver Assistance Systems, e.g. Davidse, 2006). Development of vehicle-based technologies which help this group of experienced drivers drive for longer may also be of benefit.

Conclusion

In conclusion, the most compelling contribution of this study is the link between drivers' self-reported difficulties in night-time driving, and their perceived driver skills. Young drivers with high safety skills and lower perceptual-motor skills were found to be more likely to experience difficulties while driving at night, when compared to older drivers. Driving at night can be a challenging task requiring certain skill sets. To reduce the risks and difficulties associated with night-time driving, it is important that drivers have the necessary skills and confidence in their driving and are aware of their own limitations. However, further research is needed to confirm the findings and to explore the underlying mechanisms of these difficulties.

Author contribution

İbrahim Öztürk: Methodology, Conceptualisation, Formal analysis, Writing – original draft, **Natasha Merat:** Methodology, Conceptualisation, Writing – review & editing, Supervision, Principle Investigator.

Acknowledgement

The authors would like to thank Professor Steve Fotios and Professor Richard Rowe for their contributions during the study planning process, Xian Liu and Yang Li for their help with the data collection and Anthony Horrobin, Michael Daly and Albert Solernou Crusat for development of the driving simulator scenarios.

Funding acknowledgement

The study is part of the HAROLD project (HAzards, ROad Lighting and Driving) which is supported by the Engineering and Physical Sciences Research Council (EPSRC: EP/S003576/1).

Data access statement

The data that support the findings of this study are available on request from the corresponding author (I.O., i.ozturk@leeds.ac.uk).

References

Borowsky, A., Shinar, D., & Oron-Gilad, T. (2010). Age, skill, and hazard perception in driving. Accident Analysis & Prevention, 42(4), 1240-1249.

Davidse, R. J. (2006). Older drivers and ADAS: Which systems improve road safety?. IATSS Research, 30(1), 6-20.

De Craen, S., Twisk, D. A., Hagenzieker, M. P., Elffers, H., & Brookhuis, K. A. (2011). Do young novice drivers overestimate their driving skills more than experienced drivers? Different methods lead to different conclusions. Accident Analysis & Prevention, 43(5), 1660-1665.

DeJoy, D. M. (1989). The optimism bias and traffic accident risk perception. Accident Analysis & Prevention, 21(4), 333-340.

Elander, J., West, R., & French, D. (1993). Behavioral correlates of individual differences in road-traffic crash risk: An examination of methods and findings. Psychological Bulletin, 113(2), 279.

Evans, T., Stuckey, R., & Macdonald, W. (2020). Young drivers' perceptions of risk and difficulty: Day versus night. Accident Analysis & Prevention, 147, 105753.

Evans, T., Stuckey, R., & Macdonald, W. (2022). Young drivers' perception of hazards: Variation with experience and day versus night. Transportation Research Part F: Traffic Psychology and Behaviour, 88, 258-280.

Freund, B., Colgrove, L. A., Burke, B. L., & McLeod, R. (2005). Self-rated driving performance among elderly drivers referred for driving evaluation. Accident Analysis & Prevention, 37(4), 613-618.

Gruber, N., Mosimann, U. P., Müri, R. M., & Nef, T. (2013). Vision and night driving abilities of elderly drivers. Traffic Injury Prevention, 14(5), 477-485.

Hayes, A. F. (2022). Introduction to mediation, moderation, and conditional process analysis Third Edition: A Regression-Based Approach.

Kimlin, J. A., Black, A. A., & Wood, J. M. (2020). Older drivers' self-reported vision-related night-driving difficulties and night-driving performance. Acta Ophthalmologica, 98(4), e513-e519.

Kimlin, J. A., Black, A. A., Djaja, N., & Wood, J. M. (2016). Development and validation of a vision and night driving questionnaire. Ophthalmic and Physiological Optics, 36(4), 465-476.

Lajunen, T., & Özkan, T. (2021). Driving behavior and skills. In: Vickerman, Roger (eds.) International Encyclopedia of Transportation. vol. 7 (pp. 59-64). UK: Elsevier Ltd.

Lajunen, T., & Summala, H. (1995). Driving experience, personality, and skill and safety-motive dimensions in drivers' self-assessments. Personality and Individual Differences, 19(3), 307-318.

Martinussen, L. M., Møller, M., & Prato, C. G. (2014). Assessing the relationship between the Driver Behavior Questionnaire and the Driver Skill Inventory: Revealing sub-groups of drivers. Transportation Research Part F: Traffic Psychology and Behaviour, 26, 82-91.

Martinussen, L. M., Møller, M., & Prato, C. G. (2017). Accuracy of young male drivers' self-assessments of driving skill. Transportation Research Part F: Traffic Psychology and Behaviour, 46, 228-235.

McKenna, F. P., Stanier, R. A., & Lewis, C. (1991). Factors underlying illusory self-assessment of driving skill in males and females. Accident Analysis & Prevention, 23(1), 45-52.

Morris, J. H., Sherman, J. D., & Mansfield, E. R. (1986). Failures to detect moderating effects with ordinary least squares-moderated multiple regression: Some reasons and a remedy. Psychological Bulletin, 99(2), 282.

Ostapczuk, M., Joseph, R., Pufal, J., & Musch, J. (2017). Validation of the German version of the driver skill inventory (DSI) and the driver social desirability scales (DSDS). Transportation Research Part F: Traffic Psychology and Behaviour, 45, 169-182.

Özkan, T., & Lajunen, T. (2011). Person and environment: Traffic culture. In Handbook of Traffic Psychology (pp. 179-192). Academic Press.

Öztürk, İ., Merat, N., Rowe, R., & Fotios, S. (2023). The effect of cognitive load on Detection-Response Task (DRT) performance during day- and night-time driving: A driving simulator study with young and older drivers. Manuscript submitted for publication.

Parker, D., Macdonald, L., Sutcliffe, P., & Rabbitt, P. (2001). Confidence and the older driver. Ageing & Society, 21(2), 169-182.

Regev, S., Rolison, J. J., & Moutari, S. (2018). Crash risk by driver age, gender, and time of day using a new exposure methodology. Journal of Safety Research, 66, 131–140.

White, M. J., Cunningham, L. C., & Titchener, K. (2011). Young drivers' optimism bias for accident risk and driving skill: Accountability and insight experience manipulations. Accident Analysis & Prevention, 43(4), 1309-1315.

Wood, J. M. (2020). Nighttime driving: visual, lighting and visibility challenges. Ophthalmic and Physiological Optics, 40(2), 187-201.

Xu, J., Liu, J., Sun, X., Zhang, K., Qu, W., & Ge, Y. (2018). The relationship between driving skill and driving behavior: Psychometric adaptation of the Driver Skill Inventory in China. Accident Analysis & Prevention, 120, 92-100.

Yılmaz, Ş., Arslan, B., Öztürk, İ., Özkan, Ö., Özkan, T., & Lajunen, T. (2022). Driver social desirability scale: A Turkish adaptation and examination in the driving context. Transportation Research Part F: Traffic Psychology and Behaviour, 84, 53-64.

Gender Equitable Human Factors and E-micromobility

Katie J. Parnell[1], Siobhan E. Merriman[1] & Katherine L. Plant[1]

[1]Human Factors Engineering, Transportation Research Group, Faculty of Engineering and Physical Sciences, University of Southampton, UK.

SUMMARY

Human Factors methodologies and principles can help to close the 'gender data gap' through equitable, user-centered research approaches and sociotechnical systems analysis. This paper presents research conducted into the use and uptake of electric micromobility (e-micromobility) through a gendered lens. Qualitative research is combined with systems methodologies to provide gender-equitable recommendations that highlight how this mode of travel can be more gender-equitable.

KEYWORDS

Gender Equitable Human Factors, E-micromobility, Gender data gap, Mixed methods

Introduction

The 'gender data gap' (Criado-Perez, 2020) refers to the lack of gender disaggregated data that enables the needs of males and females to be identified independently, in order to understand and develop systems that provide equitably for their differing needs (Criado-Perez, 2020). The 'gender data gap' is responsible for the design of systems, procedures, technologies and equipment that do not enable females to have equal levels of safety, opportunity or well-being in comparison to their male counter parts. This includes critical issues such as females increased injury risk when travelling in road vehicles (Linder & Svedberg, 2019) and poor fitting personal protective equipment (PPE) equipment (Fidler, 2020; Niemczyk et al, 2020).

As a discipline, Human Factors and Ergonomics (HFE) holds considerable opportunity to close the gender data gap (Read et al, 2022). HFE is defined by the International Ergonomics Society as "*the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data, and methods to design in order to optimize human well-being and overall system performance*.". When we consider the human, their interactions, well-being and impact on system performance, we must consider the individual characteristics that influence these human experiences in order to be inclusive. This is the aim of Gender Equitable Human Factors (GE-HF, Parnell et al, 2022).

The gender data gap is particularly pertinent to the transportation domain, where only 22% of workers are female (European Commission, 2021), and where these female employees are less likely to be in the higher paid decision-making roles (Department for Transport, 2020). Previous work has identified key gender equity issues within our current transport systems (Parnell et al, 2022), yet new modes of travel are emerging in the form of electric micromobility (e-micromobility). This includes small, lightweight, electric powered and personally driven transportation modes, specifically electric bicycles (e-bikes) and electric scooters (e-scooters). Although they are a relatively new mode of travel, the literature has identified that e-micromobility platforms are more likely to be used by males (Reck et al, 2021). Young, white males in particular tend to be early adopters of new technologies and are more willing to expose themselves to the higher level of risk which are characteristic of the mode. Females tend to be more safety conscious, which limits them from feeling comfortable when using micromobility due to the inadequate infrastructure for these modes of travel (Haynes et al, 2019). E-micromobility offers the opportunity to enhance model shift away from personal road vehicles, especially within more built-up urban areas, yet this will only be effective if the uptake is significant. Therefore, e-micromobility must be an attractive transport option to a diverse range of the population. Early uptake by males suggests there may be some gender factors influencing e-micromobility use.

We present work that aimed to review of the role of gender in e-micromobility transport by conducting an analysis of the motivating factors, as well as the barriers, to e-micromobility use. This work applied a combined top-down and bottom-up research approach to generate guidance that can help to ensure our future transport systems are gender-equitable. We conducted interviews and focus groups to collect user-centred qualitative data that we disaggregated by gender to identify any differing motivations and barriers to e-micromobility use. We then combined this analysis with sociotechnical systems approaches (cognitive work analysis and actor map analysis) to identify the actors and sources of responsibility in tackling gender-equity within this transportation mode. This combined approach is a valuable method for developing impactful recommendations that target gender-equity issues, something that current transport policy recommendations do not do.

Methodology

Interviews and focus groups were conducted with 24 members of the public. An equal gender split was recruited, and participants were matched on age characteristics (average age=44.33 years Range: 22-68 years, SD: 19.02 years). Online and in-person options for participation were given to enhance the inclusivity of research participation. The semi-structured interview questions aimed to obtain insight from users and non-users of e-micromobility transport on their motivations, perspectives and barriers to using both e-scooters and e-bikes. The transcripts were qualitatively analysed and deductively coded to the gender factors framework developed from more traditional transportation modes (Parnell et al, 2022). This aimed to understand how gender factors such as family roles, perceived safety, infrastructure, ergonomic design and user behaviour relate to e-micromobility travel. We combined this qualitative data with sociotechnical systems methodologies, including an abstraction hierarchy, from the cognitive work analysis tool kit, to capture the values and priorities of e-micromobility travel. An actor map analysis was also applied to identify responsible actors in the broad sociotechnical system comprising e-micromobility.

Findings

Qualitative insights into e-micromobility were reviewed with respect to key gender factors that have previously been identified within the transport domain. Combining these qualitative outputs with the wider sociotechnical systems analysis led to the generation of key recommendations that account for gender and make e-micromobility more accessible to both males and females. The recommendations target the future design and integration of e-micromobility and highlight how it can encourage more inclusive and safe use, which would in turn incentivise greater modal shift away from privately owned cars. Through the actor map analysis, key actors that hold responsibility for helping to close the data gap are also identified, enabling the recommendations to be directed to those who have the impact to inform required changes.

Conclusion

This work strives to close the 'gender-data gap' within the development of future transport modes and identifies the key role that a Human Factors approach can play in providing equitable research practises.

Acknowledgements

We are grateful for support from the Decarbonising Transport through Electrification (DTE), A whole system approach Network+ funded by EPSRC grant reference EP/S032053/1 for funding this work. In addition to this, Dr. Katie Parnell acknowledges support from University of Southampton Anniversary Fellowship programme. We would also like to give thanks to the support of the Close the Gender Data Gap group at the University of Southampton (http://closethedatagap.soton.ac.uk/).

References

Criado-Perez, C. (2019). *Invisible women: Exposing data bias in a world designed for men.* London, Penguin Random House.

Department for Transport (2020) Gender Pay Gap Report and Data https://www.gov.uk/government/publications/dft-gender-pay-gap-report-and-data-2020/dft-gender-pay-gap-report-and-data-2020 [Accessed 01/02/2023]

European Commission (2021). Women in Transport. https://transport.ec.europa.eu/transport-themes/social-issues-equality-and-attractiveness-transport-sector/equality/women-transport_en [Accessed on 01/12/2022]

Fidler, H. (2020). PPE:'one size fits all'design is a fallacy. *Nursing Standard. 35*(6), 23.

Haynes, E., Green, J., Garside, R., Kelly, M.P. and Guell, C. (2019) 'Gender and active travel: a qualitative data synthesis informed by machine learning', *International Journal of Behavioral Nutrition and Physical Activity, 16*(1), p. 135

Linder, A., & Svedberg, W. (2019a). Review of average sized male and female occupant models in European regulatory safety assessment tests and European laws: Gaps and bridging suggestions. *Accident Analysis & Prevention, 127*, 156-162.

Niemczyk, S. E., Arnold, L., & Wang, L. (2020). Incompatible functions: Problems of protection and comfort identified by female police officers required to wear ballistic vests over bras. *International Journal of Fashion Design, Technology and Education*, *13*(2), 165-172.

Parnell, K. J., Pope, K. A., Hart, S., Sturgess, E., Hayward, R., Leonard, P., & Madeira-Revell, K. (2022). 'It'sa man's world': a gender-equitable scoping review of gender, transportation, and work. *Ergonomics*, 65(11), 1-17.

Reck, D. J., & Axhausen, K. W. (2021). Who uses shared micro-mobility services? Empirical evidence from Zurich, Switzerland. *Transportation Research Part D: Transport and Environment, 94,* 102803.

H-FIT: Assessing the human factors impact of proposed changes to the railway

Nora Balfe[1]

[1]Iarnród Éireann Irish Rail

SUMMARY

This paper describes the background, development, and content of a new tool, H-FIT, to assess the likely human factors impact of proposed railway change projects. The tool provides a structured approach to identifying the scope and requirements for human factors integration at early project stages, around which the human factors activities can be specified and planned. The core of the tool is 14 design scope factors which range from planned changes to the work environment, to the introduction of new HMIs, and changes in working hours. These design scope factors are linked to physical and organisational design outcomes, such as accessibility, usability, and fatigue. Human factors goals can be set for each area of design scope against the related design outcomes.

KEYWORDS

Rail HF, Human factors integration, human performance, socio-technical system design

Introduction

Any change to railway operations can have an impact on human performance, from large changes such as the introduction of new fleet vehicles to smaller changes like the extension of a railway platform. New railway fleet vehicles, for example, should consider physical ergonomics attributes such as accessibility and comfort for staff and passengers, as well as cognitive ergonomics aspects associated with the usability of controls and displays, and organisational ergonomics aspects relating to the way information is transmitted to the vehicles for display to staff and passengers. Smaller projects may have similar considerations, but require a lower level of analysis or specific human factors expertise due to the scale and complexity of the change. A platform extension, for example, may only require a check that accessibility and information provision are unaffected for passengers, and that signal and platform-train interface visibility are unaffected for train drivers.

The inclusion of human factors (HF) in management of railway change is embedded in mandatory standards and legislation in the European rail sector, with reference to the application of human factors knowledge made in the 2016 Railway Safety Directive (EU, 2016). The EU Common Safety Method (CSM) on Safety Management Systems (SMS) is more specific and requires the integration of human factors to "address risks associated with the design and use of equipment, tasks, working conditions and organisational arrangements, taking into account human capabilities as well as limitations, and the influences on human performance" (EU, 2018, Clause 4.6.1).

Irish Rail have addressed this legislative requirement through the development of a human factors strategy, part of which is a human factors assurance (or integration) process for all proposed changes to plant, equipment, infrastructure or operations (PEIO). This sits within the wider PEIO change management and safety assurance framework under the Irish Rail SMS. When a change is proposed, the potential human factors impact is assessed, and the scope of human factors assurance activities is defined. However, it can be difficult to identify the specific human factors inputs and level of effort needed for a project at an early stage. Yet this is when the requirements are being developed for suppliers to tender against and it is necessary to provide a scope of the required HF work. This paper describes a tool, the Human-Factors Impact assessment Tool (H-FIT) developed to support the early identification of the scope of HF issues generated by the individual project. This helps to specify the correct level of human factors input in the project tender documentation.

H-FIT Development

The overall human factors assurance process at Irish Rail broadly follows the human centred design process (ISO 9241-210, 2010), but the H-FIT tool specifically focuses on the aspect of understanding and specifying the context of use and specifying high level user requirements. Existing Human Factors Integration (HFI) tools in the literature are typically based around the US Army's MANPRINT model (Houghton et al., 2015) which sets out seven domains of HFI: Staffing, Personnel, Training, Human Factors Engineering (HFE), Health Hazards, and System Safety (Widdowson & Carr, 2002). Early human factors assessments may be carried out against these domains, particularly in the defence industries (e.g., Lilliane & Jacques, 2009). However, this model is not always directly applicable to other industries. For example, in rail, the human factors discipline has relatively little input to the staff aptitudes and experience under the 'Personnel' domain since occupational psychologists have a mandate in this area. It does not therefore need to be specified under the human factors requirements for a particular project. Similarly, system safety is managed by an appointed Safety Assurance Manager as part of a mandatory safety assurance process, and while it is important that HF feed in to their analysis, that work is not specified, developed, or led by human factors. The HFE element of MANPRINT is unique to the human factors discipline, but is it still a very broad topic and it can be difficult to identify the specific and unique HF activities needed to support a particular project at early project stages. The H-FIT tool has been developed to be more specific to the HF inputs on railway change projects.

H-FIT was developed through a review of human factors taxonomies, both retrospective (supporting accident and incident investigation) and prospective (supporting human reliability analysis), including HFACS (Shappell & Wiegmann, 2000), MEDA (Rankin et al., 2000), and THERP (Swain, 1964), but particularly the 5x5 model developed by staff at the European Union Agency for Rail (ERA; Accou & Carpinelli, 2022). It was also influenced by industrial tools which the author has had the privilege to work with, but which have not been published externally. The aim of the review was to identify factors for inclusion in the H-FIT tool. The 5x5 model was a particular focus because this taxonomy is already embedded as a mandatory taxonomy in draft EU rail legislation on rail accident and incident data analysis (EU, 2020) and is proposed by ERA to be used as a non-mandatory taxonomy in the identification of human factors goals for change management, as part of forthcoming guidance which the author has contributed to.

The 5x5 model was developed by Accou & Carpinelli (2022) to provide a structured taxonomy to support the SAfety FRactal ANalysis (SAFRAN) post-incident investigation method. The aim of the SAFRAN method is to guide investigators in understanding how the composition of the safety management system may have influenced the operational decisions and actions involved in an incident or accident. The 5x5 taxonomy was initially based on a set of performance influencing factors identified in a comprehensive review of worldwide railway investigations by Kyriakidis et al., (2015). It is composed of five groups: dynamic staff, dynamic situational, static staff, static situational, and socio-interactional. The distinction between staff and situational refers to factors relating to the individual versus factors relating to the task, environment, or situation, while the static/dynamic distinction refers to the variability of the factors over time. The socio-interactional group covers teamwork and communication. Each group is composed of five factors (hence 5x5), giving 25 factors in all. The full set of 25 factors in the 5x5 tool is shown in Figure 1.

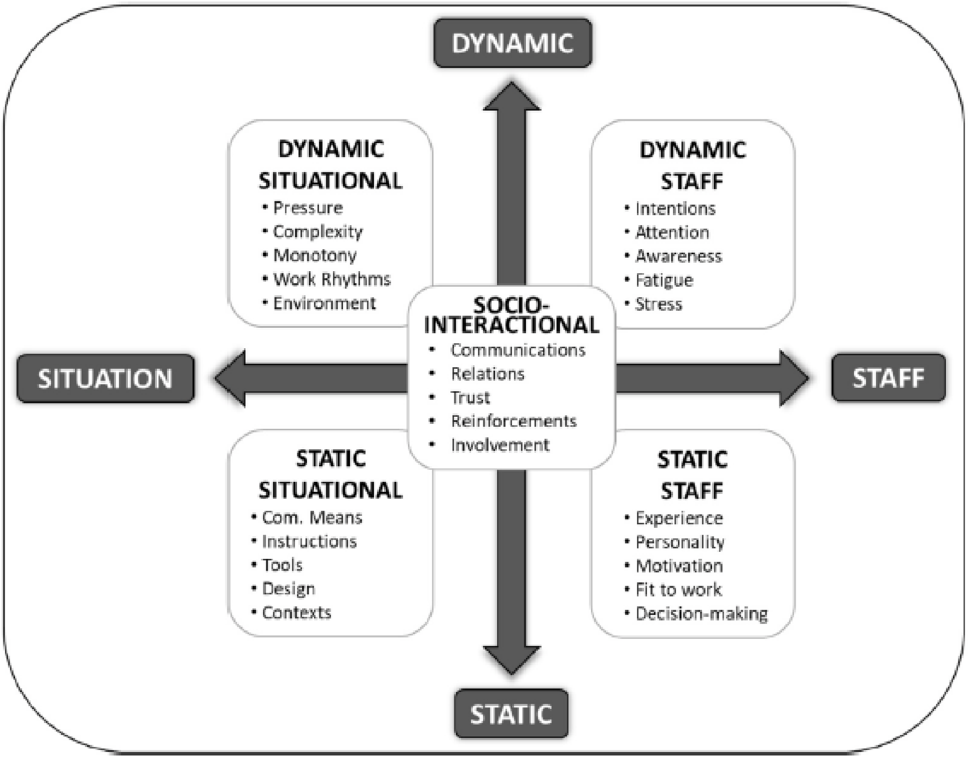


Figure 1: The 5x5 model (Accou & Carpinelli, 2022)

Within Irish Rail, the 5x5 model is not currently proposed to support human factors integration for several reasons; first, although simpler than many other human factors taxonomies, some of the factors within the model are not readily understood and some appear to overlap. For example, monotony and work rhythms have a potential overlap, as do reinforcements and motivation. Second, due to its origins as a retrospectively applied taxonomy, some of the factors are difficult to specify at early design stage. For example, the 'intentions' factor refers to the motivations for staff actions in an incident and does not apply directly at design stage. Similarly, fit to work refers to the state of a specific staff member during an incident and is not especially relevant to design. The third reason why H-FIT includes a new framework relates to the distinction between tangible, planned changes (referred to as design scope in H-FIT) and the effects of those changes on the people working in the redesigned system (referred to as design outcomes in H-FIT). In common with other human factors taxonomies, 5x5 mixes these two categories, for example communication means, instructions and tools are all tangible elements *of* the design while pressure, fatigue and stress are affected *by* the design. A clear distinction is drawn between these two elements in H-FIT with the design scope setting the required level of human factors input, and the design outcomes driving the human factors goals and activities.

Design Scope Factors

14 design factors were identified for inclusion in H-FIT (see Table 1). These are drawn from the four interaction areas within the SHELL model of human factors (Hawkins & Orlady, 1993) of Liveware-Software (L-S), Liveware-Hardware (L-H), Liveware-Environment (L-E), and Liveware-Liveware (L-L). Factor 1 in H-Fit is the environment (L-E), while factors 2-6 relate to technical changes to the system involved physical equipment, graphical user interfaces, new observable tasks, or the introduction of automation (L-H). Factors 7 and 8 relate to procedural changes (L-S), and factors 9-14 relate to more organisational factors, covered to some extent by L-L in the SHELL model, but encompassing more than just interactions with colleagues. The strong inclusion of organisational factors within H-FIT is driven by the increasing focus of European human factors on

the organisational elements, highlighted by the adoption of the term HOF, or Human and Organisational Factors, in the European legislation.

Table 1: 14 design factors

Factor	Type	Description
Factor 1	Environment	Any change to the environment where the task (including customer tasks) takes place. This may involve a change of location of the task, or changes within the location.
Factor 2	Tasks	Any change to the way safety-critical or safety-related tasks are performed, or the introduction of new safety-critical or safety-related tasks
Factor 3	Tools/equipment	Any change to existing equipment or tools used for safety critical or safety related tasks, or the introduction of new such tools or equipment
Factor 4	HMIs	Any change to existing HMIs used by safety critical staff in the course of their duties, or the introduction of a new HMI
Factor 5	Alarms	Any change to the number, format, or presentation of alarms to any role
Factor 6	Automation	Any change in the level of automation used by any role
Factor 7	Procedures	Any change to safety critical or safety related procedures, including the development and implementation of new procedures
Factor 8	Communication protocols	Any change to the communication protocols used to support safety critical communications
Factor 9	Staffing levels	Any change to the expected number of staff allocated or available to complete tasks
Factor 10	Resource availability	Any change to the availability of tools and equipment or other necessary resources
Factor 11	Roles and responsibilities	Any change to the roles and responsibilities of safety critical or safety related staff
Factor 12	Information provision	Any change to the way in which safety critical or safety related information is provided to any of the affected users
Factor 13	Leadership and supervision	Any change to the level or quality of supervision available to safety critical staff
Factor 14	Working time	Any change to the rostering of safety critical staff

There is some overlap between the factors in H-FIT, for example if Factor 2 (Tasks) is affected, it is likely that Factor 7 (Procedures) may also be affected. However, this is not necessarily the case and this means that two factors cannot be combined. For example, the format or structure of procedures may be amended without any change to the task itself. Factor 5 (Alarms) could also be regarded as a specific case of Factor 4 (HMIs), but given the critical importance of alarms and the specific requirements relating to their use, it was deemed important to have a separate category. Similarly, Factor 8 (Communications protocols) could be regarded as a sub-set of Factor 7 (Procedures), but safety critical communications are another critical topic which deserves its own category.

Design Outcomes

The design outcomes relate to the effect the design may have on human performance. In contrast to the design scope, which sets a prescriptive set of 14 factors for consideration, the listed design outcomes are intended as a set of prompts which may be supplemented from the analyst's experience. Table 2 shows the design outcomes currently included in H-FIT, mapped to the most applicable design scope factors. An argument could be made that almost all design scope factors can be related to the design outcomes; for example, a noisy environment may increase fatigue or low staffing levels could impact on the quality of teamwork. However, in practice there are some design scope factors with an obvious, direct influence on certain design outcomes. The mapping in Table 2 attempts to highlight these direct influences, and is therefore just a guide to help focus attention on the most likely areas of impact. Individual projects may have obvious direct links which do not appear in this table, and these should still be identified when analysing relevant design outcomes for the project.

Table 2: Mapping of design scope factors to design outcomes

	Design Outcomes															
	Visibility	Audibility	Thermal Comfort	Accessibility	Physical workload and stress	Mental workload and stress	Situation awareness	Usability	Human reliability and error	Fatigue	Quality of communications	Teamwork	Training needs	Motivation/job satisfaction	Risk awareness	Culture
F1: Environment	•	•	•	•	•		•		•					•	•	•
F2: Tasks					•	•			•	•			•	•		•
F3: Tools/equipment		•		•	•				•				•			
F4: HMIs	•			•		•	•	•	•				•			
F5: Alarms	•	•				•	•	•	•				•		•	
F6: Automation					•	•	•	•	•				•	•		
F7: Procedures						•	•	•	•		•	•	•		•	•
F8: Comms protocols							•		•		•	•	•			
F9: Staffing levels					•	•			•	•				•		
F10: Resources					•	•			•					•		•
F11: Roles									•		•	•		•	•	•
F12: Info provision							•		•					•		
F13: Supervision									•			•		•	•	•
F14: Working time									•	•				•		

H-FIT Structure

The overall structure of H-FIT is shown in Figure 1. The tool is held in an excel workbook which guides the user through each step. The content of the four steps, which are aligned to the ISO-9241-210 (ISO, 2010) process, are outlined below.

Figure 2: Steps in the Human Factors Impact Assessment tool for Railway Change Management

Step 1: Outline change and identify users

The first step is to briefly describe the proposed project scope and to identify the end users who may notice a change, for example, railway passengers, train drivers, signallers, etc. The purpose of this step is to capture a qualitative description of the project and to start to identify potentially affected end users. It may not be possible to provide a detailed description, but high-level information should be available from existing project documentation or discussion with the project manager. As the roles within the railway are well known and understood, a detailed target audience description is not necessary within this step.

Step 2: Identify scope of design changes

The second step is to identify the scale of the proposed change against the 14 design scope factors (see Table 1). As they are all intended to be tangible changes, they could be identified by the project manager, although in practice the table may be completed by a HF Specialist in consultation with the project manager.

Each of these 14 factors can be scored from 0 (no change) to 3 (high change). This allocation determines the level of human factors input required for a particular project. Projects with no HF impact do not require any further input; the level of effort is tailored for low-high impact projects with low impacts only needing a review of the risk assessment or a short consultation with staff for example, while high impact projects require a structured human factors integration process to plan and document the human factors activities undertaken. Some examples of the different levels for three of the factors from Table 1 are shown in Table 3. These examples are provided within the tool to provide guidance in the appropriate level to allocate for each project. The overall project is rated no, low, medium, or high human factors impact based on the highest scoring element across the 14 factors.

Table 3: Example of levels of change for three H-FIT factors

	Factor 1: Environment	Factor 6: Communication protocols	Factor 12: Working time
High (3)	A new control room or drivers cab	Introduction of new suite of forms supporting safety critical communications	Move from day to night working
Medium (2)	Installation of new lighting system within a maintenance depot	Change to a single safety critical instruction	Change from 8 to 12 hour shift pattern
Low (1)	Changed layout of a customer car park	Minor change to an existing form used to support safety critical communications	Change in time available for handovers

Step 3: Identify potential effects of design changes

The third step identifies the possible effects of a change for each affected user type against each of the factors scored above zero on step 2. The aim of this step is to start to identify the required human factors activities within the scope of the project. This step requires a much higher degree of human factors expertise, as judgements are made on the expected effect of the design scope on human performance.

For example, a new fleet of trains would represent a new working environment (amongst other factors) for train drivers (amongst other affected users) and design outcomes should be considered relating to the lighting levels within the train cab, visibility within and from the cab, noise and temperature levels within the cab, accessibility into the cab, and accessibility of the provided driver seat and console. From the passenger perspective, a new information system may be provided on-board and the visibility and usability of the information presented on that system should be considered.

Step 4: Set human factors goals

The final step is to set a human performance goal for the identified effects. Currently these are broad statements such as ‘The desk shall accommodate a user from 5PF to 95PM’ or ‘Glare shall be minimised’. Each project is then responsible for identifying the relevant standards and HF activities to achieve these goals as part of the human factors integration process.

Conclusions

The H-FIT tool presented in this paper has been developed from existing human factors taxonomies and methods to address the specific needs of Irish Rail in meeting the legislative requirements on change management, and providing sufficient information at an early project stage to support tender documentation. The tool provides a structured approach to assessing the degree of human factors impact of a proposed change, and to tailoring the planned human factors activities depending on project complexity. The objective is to set high level human factors goals or requirements before the development of a detailed human factors assurance or integration plan.

The tool has been iteratively adjusted during application against projects at Irish Rail, with changes made to the structure and guidance, and it will likely continue to evolve. A planned development is to provide more guidance on relevant standards for each design scope factor and design outcome and to identify relevant assessment methodologies.

References

Accou, B., & Carpinelli, F. (2022). Systematically investigating human and organisational factors in complex socio-technical systems by using the “Safety Fractal Analysis” Method. Applied Ergonomics, 100.

European Union (2016). Directive (EU) 2016/798 of the European Parliament and of the Council of 11 May 2016 on railway safety. Brussels: EU.

European Union (2018). Commission Delegated Regulation (EU) 2018/762 of 8 March 2018 establishing common safety methods on safety management system requirements pursuant to Directive (EU) 2016/798 of the European Parliament and of the Council and repealing Commission Regulations (EU) No 1158/2010 and (EU) No 1169/2010. Brussels: EU.

European Union (2020). Final draft of Commission Delegated Regulation (EU) establishing common safety methods for assessing the safety level and the safety performance of railway operators at national and Union level. Downloaded February 2023 from www.era.europa.eu.

Hawkins, F.H., & Orlady, H.W. (1993). Human factors in flight, 2nd ed. England: Avebury Technical.

Houghton, R., Balfe, N., & Wilson, J.R. (2015). Systems Analysis and Design. In J.R. Wilson & S. Sharples (Eds.) Evaluation of Human Work, 4th edition, pp. 221-248. Boca Raton: CRC Press.

Kyriakidis, M., Majumdar, A., Ochieng, W.Y. (2015). Data based framework to identify the most significant performance shaping factors in railway operations. Safety Science, 78, 60–76.
Lilliane, C. & Jacques, A. (2009). Body scanning technology: An early human factors analysis. In Proceedings of the 43rd Annual International Carnahan Conference on Security Technology. Zurich, Switzerland, pp. 95-99.
ISO (2010). Ergonomics of human-system interaction. Part 210: Human-centred design for interactive systems. Geneva: ISO.
Rankin, W., Hibit, R., Allen, J., & Sargent, R. (2000). Development and evaluation of the maintenance error decision aid (MEDA) process. International Journal of Industrial Ergonomics, 26(2), 261-276.
Shappell, S.A. & Wiegmann, D.A. (2000). The Human Factors Analysis and Classification System (HFACS). Washington, DC: US Department of Transportation, Office of Aviation Medicine.
Swain, A.D. (1964). Technique for Human Error Prediction (THERP). Albuquerque, New Mexico: Sandia Corp.
Widdowson, A., & Carr, D. (2002). Human Factors Integration: Implementation in the Onshore and Offshore Industries. Norwich: HMSO.

Situation Awareness in Railway PICOPs

Abigail Palmer[1] & Nora Balfe[1]

[1] Iarnród Éireann Irish Rail

Abstract

Situation awareness in the railway is extremely important due to its dynamic nature which means that operators have to perceive, understand, and act on continuously changing information. This is true in engineering work, as well as operations. During engineering works, a PICOP (person in charge of possession) needs to keep track of multiple different factors, such as the number of RRVs/OTMs being used, number of worksites, points movements, and train movements, to ensure the safe running of the requiring high levels of situation awareness. Observations were undertaken to gain an understanding of the job as a whole in order to ultimately generate recommendations on how to better support situation awareness. This paper maps the PICOP role to Endsley's three levels of situation awareness.

Factors include track layouts, and the number of worksites

KEYWORDS

PICOP, Situation Awareness, Railway

Introduction

Situation awareness is defined as the understanding an individual holds of the events occurring in their current situation (Endsley, 2021). In complex systems i.e., railways, conditions can rapidly change. Individuals need good levels of situation awareness to help make the best decision they can in any given moment. Situation awareness, alongside an organisation's rules and procedural information, allow for decision making to occur. Endsley's (2021) situation awareness model can be divided into three sublevels (Figure 1). Level one: perception, two: comprehension and three: prediction. The levels are linked but do not necessarily occur in strict chronological order. The levels represent increasingly higher levels of situation awareness. An individual with good amounts of perception and comprehension of situation awareness can combine to lead to a great amount of level three situation awareness, prediction.

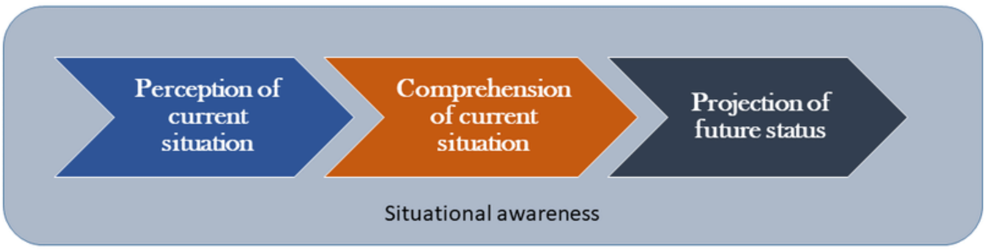


Figure 1: Three levels of Endsley's Situation Awareness (David, Lobov & Lanz 2018)

Signallers are responsible for the rail network and are in control of almost all movements on the line. When engineering or maintenance work is required on the track, or neighbouring infrastructure that cannot be completed in between the daily running schedule of trains, a T3 possession is required. A T3 possession is where a PICOP (person in charge of a possession) takes responsibility for a section of the track from a signaller in order for maintenance to occur. The PICOP is then responsible for what happens within that section of the track: they take 'possession' of the railway from the signaller. A possession often includes multiple worksites within its limits, each controlled by an ES (Engineering Supervisor). The ES controls anything that occurs within their worksite limits (Figure 2). The PICOP oversees and coordinates with ESs setting up their worksites and the two need to communicate and cooperate to control movements entering and/or exiting worksites. PICOPs are also responsible for liaising with the signaller to move points as required and ensuring OTMs (on track machines) and RRVs (road rail vehicles) have the correct route set for any movements needing to be made. The PICOP needs to give permission to operators of RRVs and OTMs to carry out any movements inside the possession but outside of worksites. There are often multiple RRVs working in one possession.

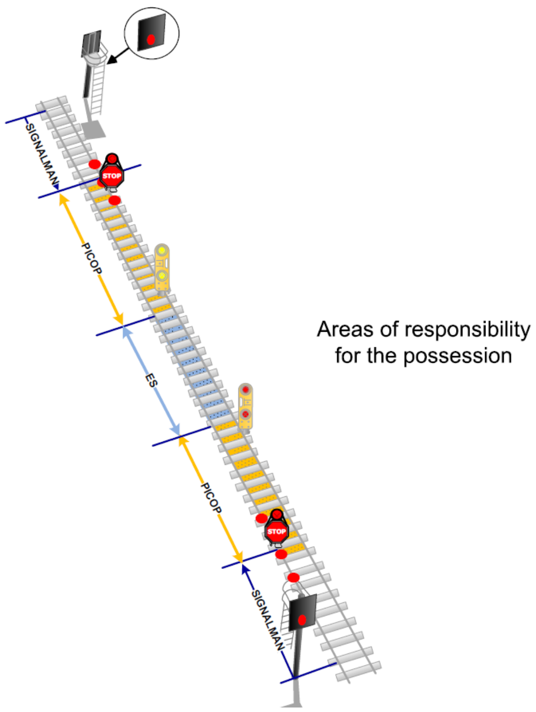


Figure 2: Areas of control within a possession (Irish Rail, 2019)

There are numerous different factors which can affect the complexity of the possession. Factors include track layouts, the number of RRVs/OTMs being used, and the number of worksites (Wilson et al., 2008). The PICOP needs to keep track of everything; therefore, it is very important that they maintain a high level of situation awareness to ensure possessions run safely and efficiently. This paper maps the three levels of situation awareness to the current provisions for a PICOP to be able to assess how their situation awareness is currently supported. With an overall project aim to highlight areas where more support may be required and provide recommendations for improvement.

Method

The main form of data collection for this project was observations. Nine observations occurred in seven different locations, observing six different PICOPs and one ES. Two observations happened during daytime possessions, while the remaining seven occurred during the nightshift, as possessions more commonly occur at night. All observations took place in the normal location of the PICOP when managing their possessions. Locations include a PICOP room (dedicated PICOP office), in the canteen, a spare office, or on the track itself. The observations were undertaken to gain an understanding of the job and what it entails, and to identify any differences in working practices. A classroom based PICOP training course was attended and semi-structured interviews with competency assessors were undertaken with the aim of understanding the training provisions in place. Semi-structure interviews with possession planners occurred to discuss and understand the planning process. The rule book and relevant procedural documentation were also reviewed to understand the relationship between the observed activities and the expected activities.

Findings

Level One Situation Awareness: What does a PICOP need to know during a possession to ensure situation awareness?

This level regards an individual's perception of their environment and everything in it (Endsley, 2021). The information can be perceived through a range of senses and can be both directly and indirectly observed. It can be through direct observation or through verbal or written communications (Endsley, 2021).

Being physically remote from the possession, a PICOP has no way of directly observing what is happening on the ground. Even when managing the possession on track, the span of the possession is usually too large for one individual to physically see all of it. PICOPs rely on pre-prepared information sheets such as work plans, possession maps alongside informational phone calls. Weekly work plans are provided which detail all the worksite information and contact details for the workers. Possession maps are static visual representations of the possession area. Each map highlights key information that is relevant to the possession, such as access points, protecting signals, required detonator positions and point numbers. All information provided is second hand and not observed by the PICOP themselves. When plant and machinery are involved in possessions, internal procedures require that controllers/PICs (Person in Charge) receive permission from the PICOP before making any movements within the possession that are not within worksite limits. Currently the PICOP uses possession arrangement forms, train forms and often a notebook in order to keep track of what is happening in the possession. A lack, or a low level, of situation awareness may cause incidents to occur which may have a financial impact to the company. An example being a points run-through, where machinery such as RRVs travel through a set of points which are not in the correct position for that movement which might cause damage the infrastructure. The damage has a cost to repair, as well as delaying the planned work during the possession.

Worksite specific details such as protecting signals and protection detonator locations as well as times when certain actions are taken, such as when the possession is granted, or when worksites have been fully set up are recorded on the forms. There is a lack of space available for all necessary information. For example, the forms have space for the locations and times points are secured at, however there is no space for noting down any point changes that occur during the possession. The PICOP either just remembers the changes or notes down the information on a scrap of paper, if available.

To help support a PICOPs level one situation awareness, it would be beneficial to redesign the possession arrangement form to improve it by providing better space for information. Endsley (1999) created a situation awareness error taxonomy to categorise the most common types of factors that affect situation awareness for each level. Factors for level one includes not having the data available, memory loss and data that is hard to detect. Memory loss and forgetting about information presented may be caused by having a high task-load which can interfere with information being attended to and distractions/disruptions, both which PICOPs can experience especially when setting up their possessions (Endsley, 1999). Providing space to write down all the information on the possession form may reduce the PICOPs reliance on memory. Golightly, Balfe, Sharples and Lowe (2012) explored situation awareness in signallers where they explained that the signalling displays allow signallers to access the information as required so that they do not have to retain the information in their working or long-time memory.

Level Two Situation Awareness: What meaning does a PICOP take from information provided?

Level two builds from the first level, as it considers the understanding an individual has of their environment and the circumstances of specific situations (Endsley, 2021). It takes all the different sources of information about an environment collected in level one and provides significance for the different elements. It can help an individual determine the seriousness of problems which may occur. Theoretically those with more experience in an area will be better at this than novices (Endsley, 2021).

To become competent as a PICOP for the first time, there is a requirement for the individual to attend classroom-based training alongside a competency assessment of them taking a possession. After becoming competent there is a refreshment cycle every two years where the PICOP needs to attend refresher training followed by a competency assessment. Some divisions organise a period of work shadowing that occurs between the classroom training and the competency assessment to allow a new PICOP to observe an actual possession being taken. This helps the new PICOP to add context to the information they received in training and see how it is applied to an actual possession.

During a possession, the PICOP relies on their local knowledge of an area in order to interpret and make sense the different forms of information provided. Local knowledge helps the PICOP know which are the best routes for RRV/OTM movements across possession and which points need to be changed to facilitate that. The local knowledge, in addition to possession arrangement forms, possession maps and any supplementary notes that have been made throughout the night, all help the PICOP to manage any movements that occur and help the PICOP to be able to deal with any information given to them verbally on the night. A lack of local knowledge may lead to mistakes such as points being moved in the wrong direction for movements which could lead to damage or derailments. The classroom-based training highly emphasises the importance of local knowledge to make sure the PICOP has context to the information they are provided, for example knowing which was to move points for specifics movements. However, there is no formal process to help PICOPs acquire this local knowledge.

Although work shadowing is organised for newly passed PICOPs in some divisions, it does not occur in all locations. The introduction of work shadowing throughout the remaining divisions may help new PICOPs in their understanding of taking possessions. Observing a possession being taken, could help new PICOPs to understand the information being given to them as it will help them take what they learnt in the classroom and see how it occurs in an authentic setting (Wilks & Ross, 2014). It can help new PICOPs understand how different pieces of information are important in the area they will be working in. It would be a great opportunity for new PICOPs to ask specific questions about why things are done in specific ways. Kopp and Minder (2016) studied job shadowing in skills trainers and found it to be a positive experience providing an opportunity to ask experts questions while they were completing the role and model good practice for the observers.

Another way to support situation awareness of PICOPs is to help ensure a PICOP has excellent local knowledge. Introducing a more formalised assessment of an individual's local knowledge would be beneficial to ensure they have the ability to assess the information provided and apply the context of the specific location to be able to help identify if/when problems occur. It is important that PICOPs start the job with excellent local knowledge of their area and maintain it throughout their career. PICOPs commonly work in one area which means PICOPs would not need to be familiar with large areas of land. Testing of local knowledge could be conducted in similar manner to that of drivers, as their route knowledge is regularly tested to ensure it is maintained.

Level Three Situation Awareness: What tools do PICOPs need to be able to anticipate potential future events?

Level three of Endsley's (2021) model of situation awareness looks to the future and the ability an individual has to predict and project into the future, to predict possible situations that might occur. Level three is achieved by combining both levels one and two (Endsley, 2021). Being able to predict possible situations aids in decision making and being proactive about choices being made (Endsley, 2021).

The railway is a dynamic workplace and PICOPs are expected to be able to deal with situations as they occur. OTMs and RRVs often have to travel through a possession to reach a specific worksite. The PICOP is responsible for planning movements and liaising with signallers to ensure points are in the correct position. RRVs often have to complete repetitive movements between locations. The PICOP will need to use their situation awareness in order to plan a route and predict any obstacles that may appear, then fix them before the movement is given permission to occur. PICOPs are also responsible for ensuring all work is completed in a timely manner to ensure that the track is clear and safe for train to run in time for the scheduled hand back. Forms and notebooks enable a PICOP to keep track of everything. They note down information such as points movements and positions to help facilitate RRV or OTM movement requests. They also encouraged the use of possession maps which give the PICOP a static view of the area.

To help support level three of Endsley's (2021) situation awareness model it may be beneficial to provide PICOPs with an interactive visual representation of the possession. Building on the current possession maps, giving PICOPs the ability to add the details of the current possession, such as worksite locations and points positions, might provide a clearer picture of what is happening on the track especially when movements are concerned. An alternative to this is allowing the PICOP to have access to a read-only version of the signalling schematic. This would allow the PICOP to see exactly what is happening in real time. The schematic might be able to help the PICOP anticipate problems or issues quicker by seeing the possession visually rather than having to read written information to determine locations and movements. Garner and Stiles (2013) explain that it is important to have up-to-date information when taking possession as it is essential for accurately understanding the environment and can help to predict potential areas of conflict.

Conclusion

A PICOP completes maintenance possessions on an almost daily basis. A high level of situation awareness is important for them to be able to manage and keep track of the different moving parts of a possession. An investigation into the role of a PICOP was conducted, with the aim of understanding what the PICOP role entails and to identify ways in which their situation awareness can be further supported.

Currently a PICOPs perception of their environment (situation awareness level one) is supported through pre-prepared documents containing information specific to that possession alongside phone calls from track workers, such as ESs and controllers of RRVs, providing information about what is occurring on the ground in the possession. Information is added to possession arrangement forms. If the information is not required on the forms, such as points movements and positions, it is either remembered or noted down on scrap paper. Redesigning the possession arrangement form to create more space for the PICOP to record additional important information would be beneficial so as to not rely on the PICOP remembering the position of assets/infrastructure memory or having to note it down on scrap paper.

Level two of situation awareness is currently supported for PICOPs through comprehensive classroom-based training where there is an emphasis on having excellent local knowledge, alongside a competency assessment before being allowed to take a possession by themselves. Local knowledge, and information documents such as possession arrangement form, work plans and possession maps all help to give the PICOP context to the information they are receiving. Work shadowing, although present in some divisions, is not offered in all divisions. A recommendation is the introduction of work shadowing in all divisions as it would enable new PICOPs to understand the significance of different pieces of information and provide examples on how possessions should be taken. Investigating the potential of developing an assessment to test a PICOPs level of local knowledge would help to ensure the PICOP has the context to identify when potential problems might occur.

Situation awareness level three where individuals predict into the short-term future, is supported through possession maps and communications between track staff and PICOPs to keep them aware of movements of RRVs being made. Potentially providing an interactive representation of the possession where a PICOP can add details and locations of specifics in their possession might help the PICOP maintain a better situation awareness of the possession as a whole. Alternatively, allowing the PICOP to access a read-only signalling schematic, where they can see everything playing out in real time could help PICOPs when identifying potential conflicts or problems.

Overall, during possessions, a PICOPs situation awareness is supported on all three levels of Endsley's model. Through the forms they complete and notes they write during a possession, to the local knowledge they need to maintain, and the documentation they are given before the possession occurs. The observations from this research have highlighted potential areas for improvement in order to provide additional support for current and new PICOPs. Such as a potential review into the possession arrangement form, assessment of level of local knowledge, work shadowing, and possible visualisation of the possession through an interactive representation of read-only signalling schematic. All of the potential areas of improvement could help increase a PICOPs situation awareness which can help with the safer and efficient running of possessions on the network.

References

David, J., Lobov, A., & Lanz, M. (2018, July). Leveraging digital twins for assisted learning of flexible manufacturing systems. In *2018 IEEE 16th International Conference on Industrial Informatics (INDIN)* (pp. 529-535). IEEE

Endsley, M.R. (1999, November). Situation Awareness and Human Error: Designing to Support Human Performance. In *Proceedings of the High Consequence Systems Surety Conference* (pp. 2-9)

Endsley, M.R. (2021). Situation Awareness, In: Salvendy, G., & Karwowski, W. (eds). *Handbook of human factors and ergonomics.* Fifth Edition. New Jersey: Wiley, 434-456.

Garner, W., & Stiles, S. (2013). Applying LEAN techniques to analyse railway possessions and isolations to identify opportunities for more effective procedures, In: Dadashi, N, Scott, A, Wilson, J. R., & Mills, A. (eds). *Rail Human Factors.* London: Taylor & Francis, pp 435-444.

Golightly, D., Balfe, N, Sharples, S., & Lowe, E (2009). Measuring situation awareness in rail signalling. *Rail Human Factors Around the World: Impacts on and of People for Successful Rail Operations, 361-369*

Irish Rail (IR) (2019). Rule Book Guidance Booklet, Person in Charge of Possession (PICOP). Dublin Irish Rail

Kopp, D., & Minder, B. (2016). Job-shadowing: Swiss health librarians observing experienced search specialists and information skills trainers in London. *Journal of the European Associated for Health Information and Libraries, 12*(1), 10-13.

Wilks, J., & Ross., K. (2014). Shadowing, "The Most Valuable Thing You Can Do". Threading Informal Classroom Experiences into Secondary Pre-Service Teacher Education. *Teacher Education Quarterly, 41*(2), 93-106.

Wilson, J. R., et al., (2008) Understanding safety and production risks in engineering planning and protection, *Ergonomics, 52*(7), 774-790. Doi: 10.1080/001400130802642211

Facilitators and barriers to a safe opioid prescribing process in general practice

Gill Gookey[1] & Mike Fray[2]

[1]East Midlands Academic Health Science Network, UK, [2]Loughborough University, UK

SUMMARY

Opioids e.g., morphine are high-risk medications that are frequently prescribed using a complex process in general practice. The current opioid prescribing process within six general practices was mapped using template analysis which highlighted high levels of variation. The Systems Engineering Initiative for Patient Safety v2.0 framework was used to identify overall aims for a safe opioid prescribing process and associated facilitators and barriers.

KEYWORDS

Opioid, Prescribing, General practice

Introduction

General practices are complex socio-technical systems that function as a subsystem of the wider healthcare system. Opioids are high-risk medicines that can create dependence and patient harm including death (Chen et al. 2019). A review (NHS England 2019) showed that in 2018, 5.6 million adults in England (13% of adults) received opioids from their General Practitioner (GP). There are high levels of variation in prescribing levels between GP's (Curtis et al. 2019) that cannot be fully explained. Weaknesses in the electronic prescribing system for controlled drugs have been highlighted as patient safety risks (Care Quality Commission 2022).

There has been no research into opioid prescribing processes in general practices in the UK but the creation of a national toolkit has been recommended to "improve the consistency of repeat prescribing processes". (Department of Health and Social Care 2021)

The use of Human Factors and Ergonomics in healthcare is highly supported (Chartered Institute of Ergonomics and Human Factors 2018) to assist with the development of solutions.

Objective

To use a Human Factors and Ergonomics approach to identify facilitators and barriers for a safe opioid prescribing process in general practice.

Methods

Six general practices were recruited. Semi-structured interviews were undertaken with up to three practice staff (clinical pharmacist, general practitioner, and administration role) per practice to understand the opioid prescribing process. Process maps for each practice were created using template analysis. Further analysis using the Systems Engineering Initiative for Patient Safety (SEIPS) v2.0 framework (Holden et al. 2013) identified overall aims for a safe opioid prescribing process and associated facilitators and barriers.

Results

17 interviews were undertaken (6 clinical pharmacists, 5 general practitioners and 6 administrative staff). The study successfully identified the key aims (figure 1) for a safe opioid prescribing process alongside the specific facilitators and barriers to achieving them (see example in table 1). Variation in processes was high between and within practices and relied heavily on the clinical system whose functionality could be enhanced. One process would not fit all practices. Improvement opportunities identified include written work procedures, clarity on roles and responsibilities, the work environment, and workload evaluation.

Conclusion

The opioid prescribing process is high-risk and complex. A safe process ensures the right patients are identified for further review and relies heavily on technology and effective communication. This work has successfully identified aims, facilitators, and barriers that can be incorporated into each individual practice to optimise system efficiency and staff well-being plus improve patient safety.

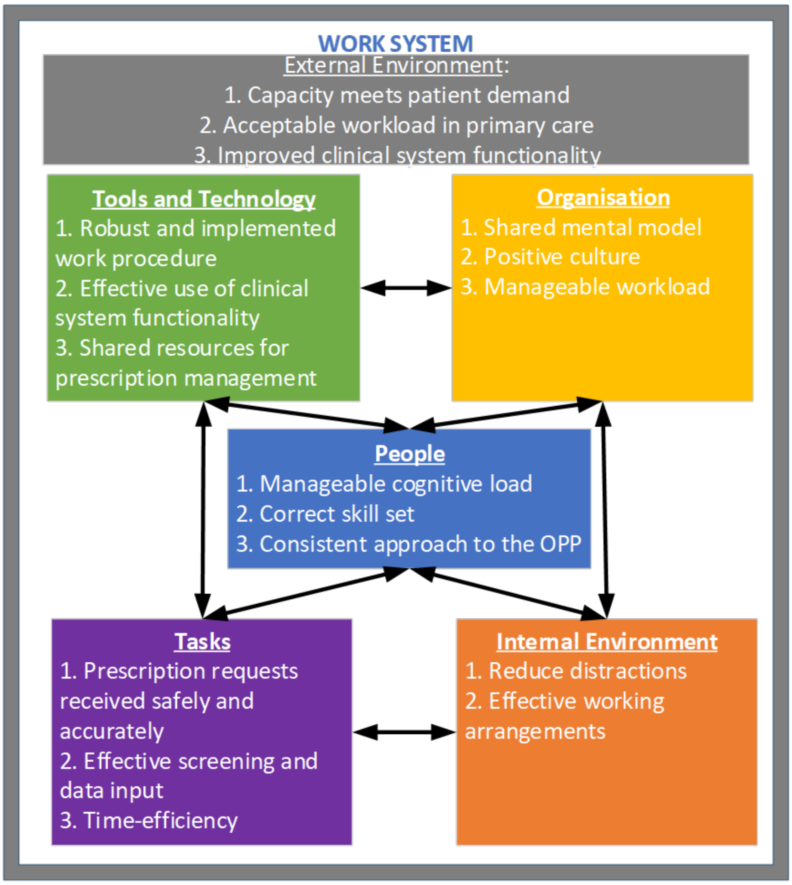


Figure 1: Work system aims for a safe opioid prescribing process using SEIPS v2.0

Table 1: An example of the facilitators and barriers identified

SEIPS area	Aim	Facilitator	Barrier
Tools and Technology	Robust and implemented work procedure	• A written, accessible, specific, and fully implemented procedure • Aligning the opioid prescription process to other high-risk medication processes	• Lack of clarity for schedule 5 opioids • Not agreeing the maximum time period between reviews for patients prescribed opioids • Written from a singular work role perspective

References

Care Quality Commission (2022), 'The safer management of controlled drugs: Annual update 2021', (cqc.org.uk: Care Quality Commission).

Chartered Institute of Ergonomics and Human Factors (2018), 'Human Factors for Health & Social Care. (white paper)', *Chartered Institute of Ergonomics and Human Factors*.

Chen, T. C., et al. (2019), 'Prescription opioids: Regional variation and socioeconomic status - evidence from primary care in England', *Int J Drug Policy,* 64, 87-94.

Curtis, H. J., et al. (2019), 'Opioid prescribing trends and geographical variation in England, 1998-2018: a retrospective database study', *Lancet Psychiatry,* 6 (2), 140-50.

Department of Health and Social Care (2021), 'Good for you, good for us, good for everybody.', (www.gov.uk).

Holden, Richard J., et al. (2013), 'SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients', *Ergonomics,* 56 (11), 1669-86.

NHS England (2019), 'The NHS Patient Safety Strategy : Safer culture, safer systems, safer patients', (NHS England: NHS England).

Safety culture in nuclear power plant construction

Teemu Reiman[1], Kaupo Viitanen[2], Jesse Hakala[3] & Karolina Wrona[4]

[1]LILIKOI, Finland, [2]VTT, Finland, [3]Oulu University, Finland [4]Dobra Consulting, Finland

SUMMARY

Safety culture related lessons learned from nuclear power plant construction projects in Finland are presented. A set of questions are proposed for organizations to discuss. The implications of these questions and constraining questions are considered.

KEYWORDS

Safety culture, nuclear safety, construction

Introduction

Strong focus on nuclear safety is required from the very beginning of a nuclear power plant's (NPP) life-cycle to avoid latent defects in the design or uncorrected errors in the construction, and to make sure the licensee develops adequate capability to safely operate the plant. To facilitate the importance of nuclear safety, organizations in the nuclear industry are required to have a good safety culture. This requirement is set in regulatory requirements and industry standards (IAEA 2016, WANO 2013). It applies to the licensee (i.e., the future operator of the plant) and other organizations participating in the design, construction, operation, or decommissioning of a nuclear power plant. International Atomic Energy Agency IAEA defines safety culture as an "assembly of characteristics and attitudes in organizations and individuals which establishes that, as an overriding priority, protection and [nuclear and radiation] safety issues receive the attention warranted by their significance" (IAEA 1991).

Culture is that which is normal to a group, and thus safety culture affects what is considered normal work, how it should be carried out, and what are the potential warning signals that would indicate risk. The main notion is that culture is something an organization creates for itself, and which, once created, influences the organization. Culture is a result of shared learning experiences that affects how the group will learn in future (Schein 2017). Safety culture can be defined as the shared values, beliefs and assumptions relating to (nuclear) safety (Reiman & Rollenhagen 2018).

Design issues have contributed to accidents across different industrial domains, with about 50 % of events in aviation, rail, chemical and nuclear industries having design errors as contributing factors (Taylor 2007, Kinnersley & Roelen 2007). Various issues connected with the design (e.g. the plant layout and siting) contributed to the Fukushima nuclear disaster during the 2011 Tōhoku earthquake and tsunami (The National Diet of Japan, 2012). In a similar fashion, decisions made in the design phase combined with good operation phase safety culture were associated with the success of the Onagawa NPP to achieve safe shutdown during the same earthquake (Reiman & Rollenhagen 2018). Quality-related problems have been noted in NPP construction projects since the 80s (NRC 1984, STUK 2006). Many of these problems have their roots in inadequate safety culture.

Design and construction of a nuclear power plant

NPP construction projects are typically carried out with a so-called EPC (Engineering, procurement, and construction) contract that is essentially a turn-key contract where the licensee buys a delivery of an operable NPP. However, a turn-key delivery does not release the licensee from its full responsibility for safety already during the construction phase. In addition to being an 'intelligent customer' during the plant delivery project, the licensee must take systematic actions to establish, foster and sustain a strong safety culture within the project supply chain. This effort is essential to ensure that all activities in the supply chain are carried out according to requirements, with quality and safety targets met to be able to achieve nuclear safety during all lifecycle phases of the NPP.

NPP projects can be considered complex adaptive systems (Reiman et al. 2015) characterized especially by interorganizational complexity (Milch & Laumann 2016). In complex systems, traditional command-and-control management style needs to be complemented with participative and distributed leadership, shared guiding principles and adaptive management (Reiman et al. 2015, Oedewald & Gotcheva 2015).

In NPP construction, complexity and safety challenges are exacerbated due to the sheer number of companies and contracts involved, long supply chains, continuously changing workforce (especially at the construction site), multiple languages and nationalities and multi-location activities. For example, approximately 2000 subcontractor companies were involved in the Olkiluoto 3 nuclear power plant construction project, reaching up to five contract tiers at the construction site. (Oedewald & Gotcheva 2015). Many of the works require specialized expertise that only few companies in the world possess, and specialized and tailored equipment is used. Quality requirements are different and typically higher than in non-nuclear construction projects. A nuclear power plant, including its systems, structures, and components, is built to withstand accident situations with very different forces and temperatures from standard operation. Many systems have a safety function in addition to their function during normal operation.

During the design phase, many important decisions are made that in addition to nuclear safety affect the reliability, industrial and radiation safety, and maintainability of the power plant. Latent weaknesses or inadequacies in design, manufacturing and construction need to be avoided by organizational processes (review, quality control, configuration management, safety analyses etc), leadership, and a healthy safety culture. With construction and decommissioning included, a nuclear power plant has over a hundred-year life-cycle. This highlights the importance of the high quality of processes (including documentation) as well as the quality of the systems, structures, and components. Finally, each NPP design needs to be licensed according to national legislation and regulatory requirements. This means that before and during the construction phase redesign of the basic design and safety analyses are carried out in addition to detailed design, construction, manufacturing, installation, and control / verification activities.

Methods and goals of the paper

There have been two major NPP construction projects in Finland during the last decade: Olkiluoto 3 NPP (OL3) and the recently terminated Hanhikivi 1 NPP (FH1). The construction of OL3 started in 2005. It is currently in test operation and planned to be in commercial operation during 2023. The FH1 licensing process, basic design and site preparatory works started in 2014, and the project was terminated during Spring 2022. We have worked as safety culture specialists and contract researchers in these two projects jointly for over 20 person years. The paper builds on our experience and lessons learned during the period from 2006-2022. A two-day lessons learned exercise was held, facilitated by a research scientist (the second author), to extract the key lessons.

We propose a set of questions that each NPP project must ask and find a shared solution, if they wish to build a culture that contributes positively to nuclear safety. We will also provide examples of constraining questions that are too narrow and easily lead to misuse the safety culture concept. We identified four categories of questions:

- defining the key concepts,
- identifying how to assess and influence culture,
- realizing what is the added value of a cultural approach to safety, and
- deciding how to consider the context.

Defining the key concepts

The first set of questions refers to the need to agree on how the definitions and models of the key concepts such as culture and safety are understood in the project. Many industries, including the nuclear, tend to use both concepts without an explicit definition (Reiman & Rollenhagen 2014). During the construction phase, this is especially confusing as nuclear safety is abstract and, on the surface, it seems to refer to an operational power plant with nuclear fuel and nuclear reaction in the core. Occupational safety issues are easier to observe, especially during construction activities. Without explicit definition, safety culture may become associated mainly with occupational issues. However, occupational safety and process, or nuclear safety are two distinct types of safety that require different approaches to manage (Hopkins 2019).

Nuclear industry has produced good guidance on the attributes of a healthy safety culture, but to fully understand what "culture" is and how "safety" is achieved during the construction phase, dialogue is needed within and between the participating organizations. More scientific models, such as Schein's (2017) model of organizational culture, clarify the essence of what culture is. Merely taking an existing definition and safety culture model and using that does not guarantee that its content is understood.

Table 1: Questions concerning the key concepts.

Question	Description	NPP construction
What safety are we talking about?	There is a need for clarity concerning what safety we are talking about: Nuclear safety, occupational safety, environmental safety, security	The distinction matters especially during NPP construction when occupational safety issues easily dominate over nuclear safety.
How do you define safety culture?	It is important to agree on a definition of safety culture that indicates how it differs from technical concepts and e.g. safety management systems.	Definition can include nuclear and occupational safety. Sometimes the term "nuclear safety culture" is used when the focus is solely on nuclear safety.
How does safety culture affect (nuclear) safety?	Since nuclear safety can be immediately endangered only after loading of the nuclear fuel, the mechanisms of influencing nuclear safety during construction need to be clarified.	The preconditions for reliable and safe operation are created during construction. Different types of organizations affect safety differently.

Defining the concepts jointly with the participating companies, and communicating them actively is important, since many organizations and individuals involved in the project lack knowledge of how safety culture relates to nuclear safety. Companies often associate safety only with occupational safety issues, not with nuclear safety. Further, it may be unclear to many how the construction phase affects the nuclear safety of an operating plant. In a nuclear power plant, structures, systems, and components may have different functions during emergency that exceed or differ from their quality requirements during normal operation. The strict quality requirements and use of certain methods and procedures, documentation requirements etc. may seem unimportant and excessive to participating companies if nuclear safety aspects are not understood (see also the question below about context). It must be constantly reminded that many of the decisions and actions made during the construction phase can have consequences years, if not decades, later. Many quality defects during construction have their roots in well-intentioned behaviour to get the job done with a quality that is "good enough" for the purpose, but without adequate understanding of what the purpose really is.

Identifying how to assess and influence culture

After defining safety and culture, the project organizations need to identify how culture can be influenced, and how one knows whether the culture supports safety or not. There needs to be an agreement on what is a nuclear safety related event during design and construction. Typical nuclear events such as a reactor scram, unavailability of safety systems, leaks, contamination, or radiation dosages cannot yet happen. What can happen are various non-conformities in design review, construction and manufacturing activities, or even in quality control and assurance (e.g. failing to notice a non-conformity). Other potential nuclear safety related events include senior management making decisions without consulting the appropriate experts, making contracts based solely on cost, or neglecting employee concerns on a safety related matter.

Table 2: Questions concerning assessment and influence.

Question	Description	NPP construction
How do we know the strengths and weaknesses in our safety culture?	This also includes the question of who defines the criteria for a good safety culture, and what model to use.	There is a need for agreement on what is a nuclear safety related event during construction and what are the warning signs of declining safety culture.
How does one assure safety culture in the supply chain?	During construction, the future operator as well as the plant supplier need to assure safety culture in their supply chain.	Complex supply chains require grading of attention to most safety-significant companies, as well as contractual requirements for safety culture.
How does one systematically influence safety culture?	How to develop safety culture? Can you influence culture directly? Leadership and systematic culture development are important.	Leadership is about creating, maintaining, and changing culture. The importance of leadership is emphasized due to the abstract nature of nuclear safety during the construction phase.

Agreement on the attributes of good safety culture needs to be achieved between the project parties. However, too simplistic models easily lead to overemphasis on easily counted and observable activities. This does not support the added value of safety culture concept and further emphasizes the tendency to associate safety culture with more easily observable occupational safety issues.

An approach to assure safety culture in the supply chain should be agreed. Contractual requirements play a major role when managing the supply chain, and the main supplier / contractor as well as the future operator of the plant need to make sure safety culture issues are included in contracts. They also need to make sure the contracts allow them access to the sub-suppliers to ensure the adequacy of their safety culture development. However, supply chain will consist of several independent companies, and it is unrealistic to assume that any contractual requirement would lead to a radical and fast culture change. Rather than aim to change the culture, the owner and the plant supplier can gradually steer the culture in the supply chain by working together with the companies on safety culture issues. Also, an awareness of how the supply chain companies really behave and what underlying cultural logics dictate their actions is important even if it is impossible to change them. It is also important to grade the attention to the most important suppliers and make sure they in turn assure safety culture in their supply chain. However, especially at the construction site contractors must be scrutinized independently of the company they are working for.

It should be realized that leaders have an essential role in the creation of culture (Schein 2017). Leadership is an important concept that needs to be properly understood by any safety critical organization. In a nuclear construction project, the importance of leadership is emphasized due to the long-time perspective needed to assure nuclear safety – values and shared priorities are essential to properly manage nuclear safety issues that, even if neglected during construction, will most likely never harm the personnel neglecting them. The line organization thus creates culture as well as safety. Safety (culture) specialists do not create the culture, they can merely facilitate and monitor its development.

Realizing what is the added value of a cultural approach to safety

The third set of questions is about realizing the added value of cultural approach to safety during NPP construction. Safety culture can be considered the link, or a moderator between the quality of the management system, its implementation, and the final product. Safety culture specialists focus on the human and organizational drivers and barriers of quality, going deeper than the quality specialists into the subjective and social issues such as norms, beliefs, and values. By adopting this wider perspective, cultural approach should contribute to systems thinking (Reiman & Rollenhagen 2014). It should also contribute to an understanding of how different types of organizations (design, manufacturing etc.) tend to have a distinct view on safety and their role in achieving it. Some of these differences are cultural, rooted in learned basic assumptions (Schein 2017).

Table 3: Questions concerning the added value of safety culture.

Question	Description	NPP construction
How does safety culture assurance differ from quality management?	Quality and safety are closely related concepts. Quality management is also an established discipline and personnel in projects recognize quality as an issue to be integrated into the management system.	Safety culture is the link between the management system, its implementation, and the final product.
What is the added value of the safety culture approach?	The project needs to understand why safety culture approach is required and what it adds to the existing approaches.	Safety culture can remind about the effects of personnel and organizations on nuclear safety during the construction phase and about the systemic influences on safety in general.

The development of a safety culture must serve the objectives of the organization, it is not a goal in itself. The development must be in line with the organization's strategy, and not, for example, a "counter-campaign" by the safety department or a program separate from the general development of the organization or the project. All project participants need to acknowledge the importance of culture for nuclear safety. Taking culture seriously has implications for the entire project governance (Oedewald & Gotcheva 2015).

Deciding how to consider the context

The fourth set of questions have to do with the context; understanding the specific requirements of nuclear construction on safety culture and deciding on an approach to systematically develop safety culture in the project.

Table 4: Questions concerning safety culture in a nuclear project.

Question	Description	NPP construction
What requirements does the context set us?	Issues to discuss include what good leadership is in this context, how the contracts facilitate / hinder good quality work and how the supply chain should be managed.	Project environment: High turnover, multicultural context with language issues, people inexperienced in nuclear, education from basic to doctorate, schedule pressures combined with heavy regulation.
How to take multicultural issues into account?	A specific question for large projects involving companies and individuals from all over the globe concerns multicultural issues.	Issues regarding leadership, authority and communication are especially relevant for the NPP construction project.
How to approach safety culture in nuclear power construction?	This is a holistic question about how to proceed with the concept of safety culture in NPP construction.	Systems thinking, shared values, understanding of nuclear safety, future orientation, information flow, and influence of organizational structural issues such as contracts are important.

A common argument is that if safety (or quality) suffers, the production will soon follow suit. The challenge is that during construction, neglect of nuclear safety may manifest years, if not decades, later. For example, design solutions influence the maintainability of the plant and the subsequent radiation dosages that the maintenance personnel will receive. Thus, the role (and instrumental value) of nuclear safety differs during the construction phase from that of the operation phase. During operation, nuclear safety is a guarantee of continued operation (of making money, that is). Even more importantly, the plant is producing electricity, and nuclear safety is a prerequisite for the continued production. However, during the construction phase, the company does not produce anything yet. During the operation phase many safety and development issues can be taken care of while the production continues and only in problematic or uncertain cases the plant needs to be shut down. Analysis and careful consideration (while the plant is in operation) are socially accepted by all parties. However, during construction, analysis and careful consideration may take time away from the schedule, and as there is no production, all issues that take time also cost money. Nevertheless, the schedule is not the enemy either, but rather an important aspect of organizing the work that should contribute to quality and safety while keeping the costs in line.

During nuclear construction, attention needs to be devoted to communication and flow of information. Networks and long supply chains naturally reduce the amount of information reaching the licensee – proactive communications and alternative channels are needed. Safety observations and safety concern reporting systems need to be set up as alternative channels for raising quality and safety issues. It must be constantly reminded that many of the decisions and actions made during the construction phase can have consequences for the safety and reliability of the operational power plant in the future. Multicultural work environment creates additional challenges for clear communication. Especially at the construction site, where multiple languages are spoken, attention needs to be paid to the clarity of messages. Translation issues, and translation errors, is another challenge. Finally, terminology differs between countries (and even organizations). A project terminology needs to be agreed with all participants. Another issue related to communication concerns the role and authority of the supervisor. In the Nordic countries communication style is rather straightforward and it is culturally acceptable for a worker to openly question management decisions. This may not be the case in some other cultures.

In the project world, the time perspective is typically short and requirements for practical achievements acute. In contrast, nuclear safety is a chronic issue that is not settled by any deadline.

The "overriding priority of nuclear safety" is much more abstract and difficult concept during construction, but even more important as a cultural guiding principle. Project environments are characterized by high turnover, multicultural context with multiple languages spoken, people with varying experience in the nuclear, varying education levels, cascading schedule pressures, and fragmentation of tasks between companies. It is a demanding environment for long-term safety thinking. Commitment and expectations of the licensee's senior management are the starting point for safety culture in the entire project. Safety genuinely needs to be a shared value in the project. A long-term development program is needed to achieve this, not only a safety assessment document for the regulator.

Constraining questions and counterproductive approaches

One of the major challenges of working with safety culture in the nuclear industry is, paradoxically, the fact that good safety culture is a (contractual, industry best practice, peer group, and regulatory) requirement. This easily leads to a situation where the supplier or a sub-supplier is trying to prove they have a good safety culture, rather than openly trying to identify their weaknesses and develop activities. This is exacerbated by the fact that there are often quite detailed behavioural and attitudinal requirements for safety culture. These requirements are then submitted by the owner to the supplier for implementation. Safety culture risks becoming an intellectual exercise, a camouflage for real operations, where the supplier knows what to say and present to the auditor or representative of the owner. Safety culture policies and programs become "fantasy documents" (Clarke 2001) and safety specialists create a new discourse, a fantasy discourse, that is practiced between safety specialists of the companies. The challenge of two worlds, one of paper and one of practice, is typical to any safety-critical organization, but it is especially prevalent in any project setting with contractual requirements for plans, programs, designs, verification reports, etc.

Table 5: Constraining questions in safety culture development.

Constraining question	Description	NPP construction
How do you quantify safety culture?	Safety analyses require that risks are quantified. If safety culture influences nuclear safety, this effect needs to be a number, or so the logic goes.	The effect of safety culture on nuclear safety may be described in the Preliminary Safety Analysis Report, but its quantification should be avoided.
What exact behaviour and actions do you want to see?	This question is often asked if safety culture is a regulatory or contractual requirement.	It leads easily to overemphasis on easily counted and observable activities, and thus a focus on occupational safety.
How do you certify safety culture or verify in audit?	Whenever there are contractual or regulatory requirements, these need to be verified in some manner, typically by an inspection or an audit.	The limits of the traditional audit approach need to be realized. Audits can reveal a lot about how the organization develops culture, but not much about culture as such.
How many cultures do we need?	Safety culture is sometimes considered as one of several cultures that an organization has or needs.	Project schedules and costs create pressure to devise a "project culture" as a counterforce to safety culture. This hinders attempts to create a company-wide culture.

Another counterproductive approach is to focus merely on easily implemented and measured activities, such as trainings and issuing posters and booklets. If these are the only activities the company does to develop safety culture, they will further contribute to the fantasy nature of safety culture. These activities are easy to verify, but they do not reveal much about the actual culture.

Senior management may sometimes think they need a “project culture” in addition to, or even to counteract, safety culture. However, this approach risks making both “cultures” superimposed and artificial. The organization should develop one culture that emphasizes safety and acknowledges the other important goals of the project, quality, schedule, and cost.

Recommendations and conclusions

Understanding the concepts of culture and safety are key to assuring nuclear safety during the construction phase. Systems thinking and a future orientation are needed to be able to consider effects that manifest in time. Values and leadership are critical for assuring nuclear safety during the pre-operational phase – the moral dimension of culture is emphasized (Rollenhagen 2010) in addition to the structural aspects (contracts, the management system) supporting its development.

For a company that is building a nuclear power plant, nuclear safety refers to how the company develops long-term organizational conditions and ability to assure nuclear safety during the entire life-cycle of the plant and how it verifies the designed and built safety conditions and ability of the plant (as designed, as built, and as documented). Safety culture in such a company should facilitate the understanding and management of the proper organizational condition and ability. Safety culture should also facilitate the discovery of any underlying weakness in the system. There are always defects, errors and mistakes in complex projects, but without trust and open climate these are not reported and may remain latent until the operation phase.

Nuclear safety, and thus safety culture, look different from the perspective of the various participating organization in an NPP project (construction, manufacturing, design, etc.). Rather than talking about safety culture idiosyncrasies of a lifecycle (e.g. construction phase) in general, it is recommended to talk about idiosyncrasies of different types of organizations. One reason for this is the fact that life-cycle phases coincide: while (non-nuclear) construction starts, nuclear related design and licensing is still ongoing, and manufacturing of so called long-lead items (such as the reactor pressure vessel) starts early and lasts well into the nuclear construction phase. Future research should clarify the differences in safety culture between the various types of organizations participating in complex safety-critical projects.

To properly consider safety culture in a project environment, contracts and supply chain management in general is in a key role: Conditions for good safety culture are created before the project execution fully starts. However, contracts are also one of the main potential hindrances to the development of safety culture since contractual arrangements may promote the above illustrated constraining questions rather than an open dialogue on cultural aspects of safety. If the contracts reward timely delivery and costs over quality, thorough verification and documentation and joint learning, the supply chain safety culture will learn and develop accordingly.

In conclusion, systems thinking, shared values, understanding of nuclear safety, future orientation, information flow, and influence of organizational structural issues such as contracts are important issues to consider in NPP construction projects. Leadership, communication, and authority issues need also to be considered especially at the multicultural construction site. The project parties need to agree on how definitions and models of the key concepts such as culture and safety are understood, and how culture can be influenced, and how one knows whether the culture supports safety or not. In addition, the added value of cultural approach to safety during NPP construction needs to be realized. Finally, an approach fit to the project realities need to be decided and implemented. Simplification and quantification of safety culture should be avoided, as well as the reduction of culture to mere safety behavior. Culture needs to be understood as result of joint learning of the group, that influences how the group perceives, thinks, feels, and acts.

References

Clarke, L. (2001). Mission Improbable: Using Fantasy Documents to Tame Disaster. University of Chicago Press.

Hopkins, A. (2019). Organising for Safety. How Structure Creates Culture. Wolters Kluwer.

IAEA. (1991). Safety Culture. Safety Series No. 75-INSAG-4. Vienna: International Atomic Energy Agency.

IAEA. (2016). Leadership and Management for Safety. No. GSR Part 2. Vienna: International Atomic Energy Agency.

Kinnersley, S. & Roelen, A. (2007). The contribution of design to accidents. Safety Science 45, 31–60.

Milch, V. & Laumann, K. (2016). Interorganizational complexity and organizational accident risk: A literature review. Safety Science 82, 9-17.

NRC. (1984). Improving Quality and the Assurance of Quality in the Design and Construction of Nuclear Power Plants: A Report to Congress (NUREG-1055). U.S. Nuclear Regulatory Commission, Washington, D.C.

The National Diet of Japan. (2012). The official report of the Fukushima nuclear accident independent investigation Commission. Tokyo, Japan.

Oedewald, P., & Gotcheva, N. (2015). Safety culture and subcontractor network governance in a complex safety critical project. Reliability Engineering and System Safety 141, 106-114.

Reiman, T. & Rollenhagen, C. (2014). Does the concept of safety culture help or hinder systems thinking in safety? Accident Analysis and Prevention 68, 5-15.

Reiman, T. & Rollenhagen, C. (2018). Safety culture. In N.Möller, S.O. Hansson, J-E. Holmberg, C. Rollenhagen (Eds.), Handbook of Safety Principles. Hoboken: John Wiley & Sons.

Reiman, T., Rollenhagen, C., Pietikäinen, E. & Heikkilä, J. (2015). Principles of adaptive management in complex safety critical organizations. Safety Science 71, 80-92.

Rollenhagen, C. (2010). Can focus on safety culture become an excuse for not rethinking design of technology? Safety Science 48, 268–278.

Schein, E.H. (2017). Organizational culture and leadership. 5th Edition. Wiley.

STUK. (2006). Management of Safety Requirements in Subcontracting During the Olkiluoto 3 Nuclear Power Plant Construction Phase. Investigation Report 1/06, 2006. Säteilyturvakeskus, Helsinki, Finland.

Taylor, J.R. (2007) Statistics of design error in the process industries. Safety Science 45, 61-73

WANO. (2013). PL 2013-01 Traits of a Healthy Safety Culture. World Association of Nuclear Operators.

Clinician perspectives around automating the Emergency Department triage process

Katherine L. Plant[1], Beverley Townsend[2], & OlTunde Ashaolu[3]

[1]Human Factors Engineering, Transportation Research Group, Faculty of Engineering and Physical Sciences, University of Southampton, UK, [2]York Law School, University of York, UK, [3]York and Scarborough Teaching Hospitals NHS Foundation Trust, UK

SUMMARY

Healthcare has arguably been the sector most impacted by the Covid-19 pandemic, leaving Emergency Department (ED) medical teams overworked and understaffed. An automated system for ED triage has been developed to help alleviate some of these pressures. Eight ED clinicians were interviewed to capture their views of the automated system. Insights were generated around where this system might add value and areas of challenge or concern. These findings will be used to refine the prototype for end-user testing and support the development of training material for clinicians.

KEYWORDS

Autonomous systems, healthcare, Emergency Department triage

Introduction

In July 2022 the Royal College of Emergency Medicine conducted a snapshot survey which showed that 94% of (87) Clinical Leads in UK Emergency Departments (ED) were not confident that their organisation would safely be able to manage forthcoming winter pressures ("RCEM snapshot survey", 2022). By October 2022, the British Medical Association reported that there were over 2.2 million ED attendances, with waiting times at record highs. The number of patients waiting over 12 hours from 'decision to admission' was 60 times higher than it was in October 2019 (impacting over 40,000 patients in October 2022) and believed to be an underestimate of actual waiting times ("NHS backlog data analysis", 2022). Healthcare has arguably been the sector most impacted by the Covid-19 pandemic, leaving ED medical teams overworked and understaffed. There is the potential for Artificial Intelligence (AI)-assisted diagnosis to alleviate some of these difficulties. The Diagnostic AI System for Robot-Assisted ED Triage (DAISY) project, funded by the Trustworthy Autonomous Systems Hub, is developing a robot system to automate the ED triage process. DAISY will enable a patient to input subjective information about their condition and will support the patient in capturing objective vital signs (e.g., blood pressure and temperature). DAISY will use its underpinning algorithms to perform an assessment with appropriate advisory information regarding a preliminary diagnosis and treatment plan, which the clinician will review and discuss with the patient. The aim of this study was to capture clinician perspectives around the introduction of this technology, to incorporate clinician requirements into future iterations of DAISY and to ensure communications and/or training content are aligned with concerns.

Method

An interview schedule, drawing on questions from the Schema Action World Research Method (Plant & Stanton, 2016; Parnell et al., 2022) and on the work describing social, legal, ethical, empathetic, and cultural (SLEEC) norms and concerns in autonomous-agent contexts (Townsend et al., 2022) was developed to capture clinician perspectives for automating the ED triage process (i.e., the DAISY system). Participants were asked to describe the current (typical) process of ED triage, including clinical decision points. Questions centred around the role of past experiences and expectations, sources of information, cultural considerations, and empathetic practice. Subsequently, participants were introduced to the functionality of the DAISY system and questions covered areas including its potential utility, influences on trust, and the role of intuition and non-verbal cues (NVC) in patient-clinician interactions. Participants consisted of eight ED clinicians representing a range of roles including Nurse Practitioner (1), Junior Doctors (3) and Consultants (4). Clinicians worked at a variety of hospitals across England and the sample consisted of a 55% male to 45% female split, noting that all of the Consultants were male. Interviews were conducted on either MS Teams or Zoom and lasted ~45 minutes. The interviews were automatically transcribed using the MS Teams function and then 'cleaned' by one of the authors for sense checking. The data were thematically analysed using both inductive (generating insights from the data) and deductive (exploring data with SLEEC norms) approaches, the former will be discussed below.

Preliminary Results

The inductive analysis grouped the data into five core themes: insights into current practice, challenges/concerns, trust and ethics, added value/benefit, and future considerations. In relation to current practice, variability in the ED triage process between hospitals was apparent which would need to be considered in a wider rollout. All clinicians, regardless of job role, stated the importance of NVC and intuition when treating patients and the role of local knowledge (e.g., known drug offenders). Concern was expressed as to how these would be accounted for by the DAISY system. Patients seeking reassurance was stated at the primary outcome of patient-clinician interactions and the DAISY system was seen as potentially advantageous in this area by freeing up clinician time from routine tasks to spend with patients. Other added values included standardising the quality of triage reports and there were interesting insights into whether patients would be more likely to disclose sources of (sensitive) injury or domestic violence to a technological agent.

Conclusions and Future Work

Clinician insights have proved invaluable at understanding end-user perspectives, which will be used to refine the prototype system for pilot testing in a custom built testbed. A secondary outcome has been to demonstrate the importance of user-centred design to the non-Human Factors members of the team (i.e., software engineers). An interactive online survey (using videos of the working prototype) is in development to capture patient perspectives of the system.

References

NHS Backlog Data Analysis (2022). Retrieved from: NHS backlog data analysis (bma.org.uk) [28/11/2022]

Parnell, K. J., Fischer, J. E., Clark, J. R., Bodenmann, A., Galvez Trigo, M. J., Brito, M. P., ... & Ramchurn, S. D. (2022). Trustworthy UAV relationships: Applying the Schema Action World taxonomy to UAVs and UAV swarm operations. *International Journal of Human–Computer Interaction*, 1-17

Plant, K.L. & Stanton, N.A. 2016. Development of the Schema Action World Research Method (SWARM) for the elicitation of perceptual cycle data. *Theoretical Issues in Ergonomics Science*, 17(4), 376-401.

Royal College of Emergency Medicine Snapshot survey and results (covering the period 27 June to 3 July 2022). Retried from: Which region is your organisation in? (rcem.ac.uk) [28/11/2022]

Identifying Non-Technical Skills behavioural markers in rail controller and maintenance roles

Anisha Tailor, Paul Leach, Kirsten Huysamen & Priya Shah

Rail Safety and Standards Board

SUMMARY

This paper discusses the use of task observations, semi-structured interviews and thematic analysis to identify and classify Non-Technical Skills (NTS) behavioural data in rail controller and rolling stock maintenance roles. A human-centred design process was applied to develop materials to support rail organisations with observing, evaluating and developing NTS in rail controller and rolling stock maintenance roles. Wider factors influencing human performance in these roles were also identified and are discussed in this paper.

KEYWORDS

Non-technical skills; competence; rail

Introduction

Non-technical skills (NTS) are "the cognitive, social, and personal resource skills that complement technical skills, and contribute to safe and efficient task performance" and have been investigated in a range of high-risk industries (Flin, 2008). Research by RSSB produced an NTS framework and an accompanying set of behavioural markers for the train driver role (RSSB, 2012). However, a rail industry consultation found that NTS integration is varied across organisations and across rail safety-critical roles, partly due to a limited understanding of what NTS look like in non-driver roles (RSSB, 2022). The objectives of this research were to identify and classify NTS behaviours and strategies for rail controllers and rolling stock maintenance staff, and to develop human-centred NTS materials to support the observation, development and measurement of NTS in these roles.

Method

Data collection

The rail industry consultation indicated gaps in knowledge about how NTS are applied in practice in non-driver roles (RSSB, 2022), so this research sought to examine a range of tasks in the field to observe and identify NTS behaviours and strategies for rail controllers and rolling stock maintenance staff. This approach is slightly different to that of a typical Risk-Based Training Needs Analysis (RBTNA). For example, the RSSB RBTNA is used, in part, to deconstruct all tasks undertaken in a role and then map NTS – taken from the RSSB NTS framework – to these tasks. This is essentially about mapping a pre-existing NTS framework to tasks, whereas the current research looked to uncover NTS behaviours and strategies that could then be used to refine and enhance the existing NTS framework. As such. a selection of tasks was instead identified with the focus on uncovering NTS behaviours and strategies as opposed to mapping existing ones to a given task. The tasks in scope for each role were identified and prioritised via task analyses. Existing RBTNAs produced by participating organisations were also used to determine the tasks in scope, to help ensure a range of behaviours could be discovered.

Site visits were undertaken at three railway control rooms, covering a passenger train operator, a freight train operator, and Network Rail. Site visits were undertaken at four rolling stock maintenance depots, covering a passenger train operator, two freight operating companies, a rolling stock manufacturer.

The researchers captured behavioural data while observing front line staff undertaking tasks (n=8 for control; n=10 for maintenance). Observations were structured around the Human Information Processing Model. Semi-structured interviews were carried out with subject matter experts, using critical incident technique questions (front line staff, assessors, instructors and front-line managers; n=13 for control; n=15 for maintenance). The interview schedule and observations were designed to elicit information on *how* staff carry out their tasks and to identify behaviours and strategies that help them perform tasks well.

Data analysis

Thematic analysis was used to code and combine the behavioural data, creating sets of behavioural statements for each of the roles. These statements were then categorised and refined using the RSSB NTS framework of seven NTS categories and 26 skills as well as the behavioural markers developed for the train driver role (RSSB, 2012).

Design of NTS materials

The NTS data were used to create resources to support the observation, measurement and development of NTS in rail controller and rolling stock maintenance staff. Structured questions were used to gather feedback from intended end users and refine the materials (front line staff, trainers, assessors and operational front line managers; n=6 for control; n=4 for maintenance).

The NTS materials

For each role, three resources were produced to support the integration of NTS.

NTS task examples: For each of the 26 skills, this provides a non-exhaustive list of example behaviours and associated front line tasks that demonstrate these skills. It illustrates what NTS looks like when completing rail controller and rolling stock maintenance staff tasks.

NTS behavioural descriptions: For each of the 26 skills, this provides a description of positive NTS behaviours that rail controller and rolling stock maintenance staff demonstrate to support safe and efficient performance. It is design to support organisations in integrating NTS into competence management systems, including training and assessments.

NTS strategies: This presents specific techniques that rail controller and rolling stock maintenance staff use to apply NTS. These aim to help front line staff understand and choose practical things they can do to apply NTS, and help managers to support front line staff NTS development.

Wider Human Factors considerations

Data collected also identified organisation and job/workplace factors which affected human performance in rail controller and rolling stock maintenance staff. In maintenance, enablers included: using task or job rotation to reduce the repetitiveness of the work (which can lead to complacency or things being missed) and allocating sufficient time to the completion of maintenance tasks to encourage staff not to rush tasks. In control, enablers included: using phones with instant replay to allow any missed details to be picked up after a call has ended and upskilling controllers on different routes so they can support each other when workload becomes too high. Rail organisations should address such factors when seeking to improve human performance, and not focus on NTS alone.

References

Flin, R. H., Connor, P., & Crichton, M. (2008). Safety at the sharp end: a guide to non-technical skills. Aldershot: Ashgate.

RSSB (2012). Non-technical skills for rail: A list of skills and behavioural markers for drivers, with guidance notes (T869). Retrieved 28th November, 2022, from: https://www.sparkrail.org/Lists/Records/DispForm.aspx?ID=10025

RSSB (2022). Enhancing the integration of non-technical skills in competence management systems: Literature review and interviews (T1207). Retrieved 28th November 2022, from: https://www.rssb.co.uk/en/safety-and-health/improving-safety-health-and-wellbeing/understanding-human-factors/non-technical-skills/introduction-to-non-technical-skills

User centred approach to developing a Concept of Operations

Kate Shield & Elaine Thompson

Arup, UK.

SUMMARY

The paper describes a user centred approach to the development of a Concept of Operations and its application to the Synthetic Environment, which is a digital signalling design tool that will be used to develop seamlessly integrated scheme designs for Network Rail.

KEYWORDS

Concept of Operations, Use Cases, User Needs, Digital Solutions

Introduction

There are many examples of systems being procured, or infrastructure designed and built without really considering how it is going to be used, the operational or task goals that need to be achieved, or how maintained. This results in a design or system where the users must 'fit' into the system, and this failure to consider the usability and operation fully can result in re-design, higher operational costs long term, dissatisfaction or reduced levels of safety as operators have to cope with a system or environment that doesn't really meet their goals. Development of a detailed and comprehensive Concept of Operations that places the users at the heart of the design can produce real benefit in ensuring that a full suite of user and system requirements are developed.

What is a 'Concept of Operations' Document?

The Concept of Operations (ConOps) is intended to be a user-oriented document that describes the characteristics of the desired future-state of a system from the viewpoint of all stakeholders, including users, developers, maintainers, passengers, business leaders and any other affected parties. The Concept of Operations document aims to:

- Serve as a tool to engage stakeholders;
- Identify the users of the proposed management system;
- Describe the context of the proposed system;
- Provide a view of how the proposed system will function through scenarios/user stories;
- Explore new ways of working;
- Support the integration of new technologies;
- Provide high level descriptions of the activity;
- Operational requirements;
- Descriptions of functionality of area / process.

There are many benefits to a Concept of Operations document, and it is a core Systems Engineering activity. Development of the document provides an opportunity to benchmark design with current and future industry best practice, creating a structured plan for how the system will ideally operate, and supports the development of the 'Basis of Design'.

There is tendency with Concept of Operations documents to focus on either the operational, infrastructure or technology changes, and not consider or engage with the end-users, or focus on only the operational element with no concern for maintenance or customers. This can result in a disjointed ConOps which is based on ‘work as imagined’ rather than ‘work as done’.

Systems Engineering and Human Factors are complementary disciplines; Human Factors can provide the user perspective for a system or operation, creating user requirements that informs system and safety requirements. By bringing these together into one unified set of requirements, the usability of the system can be fully integrated throughout design development through to testing, and commissioning into use.

Developing a User Centred Concept of Operation

The Concept of Operations should be utilised as a living iterative document that is used to engage the end-users from the start till the end of any project rather than just a designers document. The steps towards creating an effective document are described below:

1. **Personas:**

Personas provide meaningful archetypes that can be used to describe user interaction with systems or environments that can be referred to at all stages of service innovation, development, and delivery. They represent significant groups of end-users and their different goals, needs, characteristics, and expectations. The focus on different user viewpoints can help inform decision-making in the design of services and the user experience. They also support the development of an understanding of the roles and functions that will interact with the system, the inputs and outputs, and key challenges. Current pain points and opportunities for efficiency can be captured at each stage in the design e.g., difficulties with design integration, points at which re-work normally occurs; opportunities to capture design information for other purposes or streamlining processes and checking. It is key to do proper research, gathering information about the current or intended users. Data collection methods include interviews or workshops with end-users and user-facing roles.

2. **Operational Scenarios or Use Cases:**

The Concept of Operations should be based on Operational Scenarios or Use Cases, (depending on which is most relevant to system that is being described) to define what the envisioned system provides or how it functions throughout the timeline of a scenario or design stages. These should include the range of scenarios that the system can be used within, including normal or optimal performance, degraded and emergency modes. The Use Cases should describe all of the modes or configurations that the system may need to have in order to deliver its intended purpose throughout its life cycle. This may include modes needed in the development of the system, e.g. testing or training, as well as operational modes and system decommissioning [Ref 1].

3. **User Stories:**

User stories are stated ideas of requirements that express what users need to perform their specific goals within the Operational Scenario or Use Case. User stories are brief, with each element often containing fewer than 10 or 15 words each. User stories are "to-do" lists that help determine the user needs for a project or system, and support their translation to user and system requirements.

4. **User Story Boards:**

User Story Boards or 'stakeholder explainers' - which can be either static and interactive visualisations bring the concept of the system or product to life. They can incorporate key elements from the ConOps work including the User Stories and identified scenarios to explain to the wider stakeholder what a system or product may look like and how the users will interact with along the timeline of the Operational Scenario or Use Case.

5. **Requirements Development:**

If the Concept of Operations is developed effectively, then the system and user requirements for the system or project should be clearly identifiable and easily extracted to support design development, or definition of operations and performance requirements. By taking the user centred approach in the Concept of Operations document, the requirements should retain the user needs at their core and result in systems and infrastructure that is easy to use, fit for purpose and can be operated and maintained safely while meeting the needs of customers.

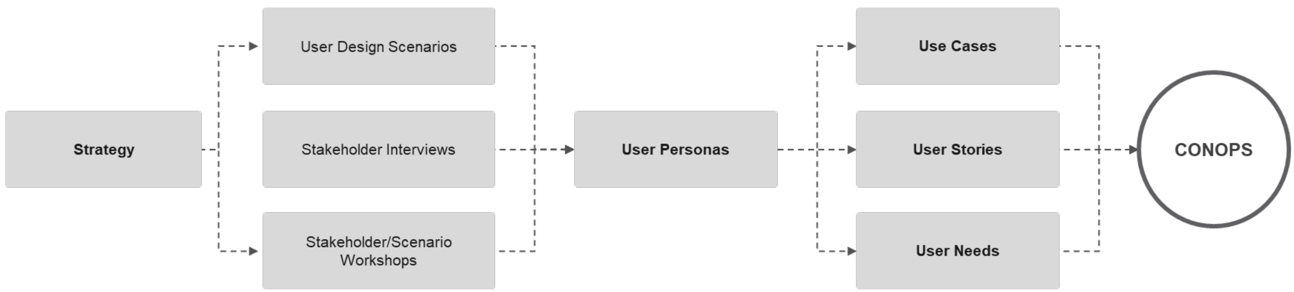


Figure 1 Illustration of a user centred approach to Concept of Operations Development

Application of the user centred approach

A user centred approach was taken to develop the Concept of Operation for a new digital tool that will allow designers to develop and deliver fully integrated, tested designs for digital signalling such as European Train Control System (ETCS). These digital components are collectively termed Future Command Control and Signalling (F-CCS) solutions. These future Digital Signalling solutions will replace traditional lineside signalling with movement authorities and speed information that is relayed to the driver via an in-cab display. The data driven design approach offers the ability for design standardisation to bring efficiency and reduce the overall cost of replacing traditional signal design with F-CCS solutions.

Network Rail are developing a strategy for the Design and Validation of F-CCS designs, that will feature a set of digital design tools at its heart. The central design tool, called the Synthetic Environment, forms a central component for a Network Rail strategy for achieving cost reductions in the design and delivery of future digital signalling schemes by harnessing technology solutions that supports fully integrated, consistent designs [Ref 2]. Synthetic Environments are computer simulations that provide a high level of realism to a physical environment, and are developed to support design and testing. Synthetic Environments are not a new concept, they are widely used in other sectors and industries [Ref 3], but they are new to the Rail Industry in the UK.

Where a complex technical solution is needed, there is a tendency to focus on the digital tool, delivering the functional capability of the tool, rather than resulting in a solution that meets end users' needs. However, to achieve a result that meets the needs of a range of users, the client realised that they needed to put the user perspective foremost. By placing the user goals and needs at the heart of the Concept of Operation, it provided a basis for developing the requirements for the solution, with the intention to create an output that is intuitive and fully supports the design process.

Concept of Operations development

There are a range of engineers, designers and end user representatives that would need to interact with the Synthetic Environment to create an F-CCS design. As this is a new concept for the Signalling design community, it was important to establish how designs are developed using current systems and processes, and how these will change for the future i.e., establish the ‘as is’ and the ‘to be’ for the signalling design process.

Stakeholder workshops were held with representatives from a range of technical disciplines who are involved in developing conventional and digital signalling designs. These workshops mapped out the range of users involved in design and assurance at each stage of the design process. This mapping included identifying the user goals and information needs using the current design tools and processes, and the future vision for using the Synthetic Environment.

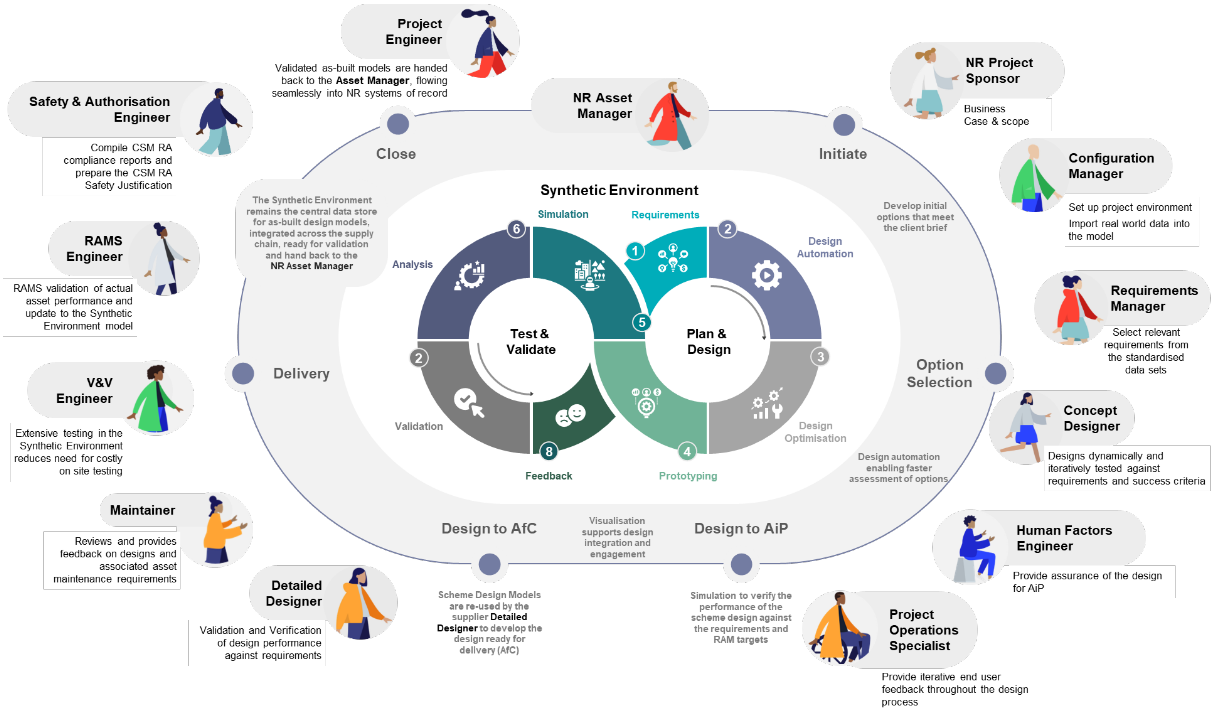


Figure 2 User interaction with the Synthetic Environment at each stage of the project delivery lifecycle

The goals and user need for the primary users of the tool were expanded to create a detailed view of user interaction at each stage of design. A set of Use Cases provided the mechanism to convey the complexity of multiple users interacting with the Synthetic Environment in parallel to perform specific tasks in order to achieve detailed design goals. In conventional signalling designs, multiple designers or disciplines would be working simultaneously on design activities before integrating the designs together. The analysis identified that the conventional approach to design was error prone due to the reliance on these design integration activities to identify any inconsistencies or lack of compatibility between disciplines. Therefore, the benefits of the Synthetic Environment to provide a single view of the design that highlights inconsistencies could be stated as part of the Concept of Operations.

The primary set of Use Cases for each stage of the design were developed, describing the user's interaction with the tool, and how the interaction changes as the design develops through the lifecycle. The Use Cases provided a basis to work with Digital Developers to develop a System Architecture that could support the functional goals for the tool.

End User Engagement

One of the limitations of the current signalling design process is the limitations for engaging with end users such as Train Drivers, Maintainers and Signallers. The conventional signalling design engagement is delivered via the Signal Sighting process, which relies on the expert user groups interpretation of signalling scheme drawings to identify any potential clashes for signal sighting.

ETCS has been delivered across a relatively limited number of schemes such as Cambrian Line, Thameslink, West Coast, Elizabeth Line, and in most cases has been delivered as ETCS Level 2 Overlay where ETCS is provided alongside conventional lineside signals. Future ETCS schemes will replace conventional lineside signals, and therefore there is an increased need to review and validate ETCS designs. In ETCS schemes, the transition between conventional and ETCS has been identified as an area of potentially increased workload for Train Drivers [Ref 4], and therefore early engagement with end users is critical to ensuring that schemes are developed to reduce the potential for overload.

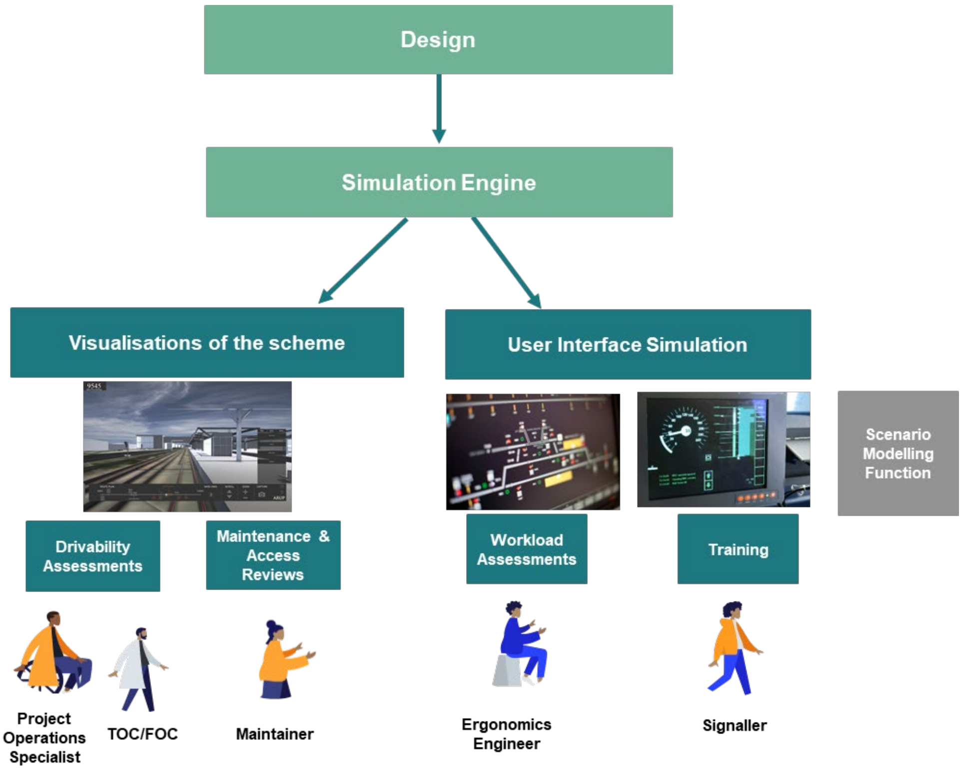


Figure 3 Examples of how visualisations and simulation will engage end users to capture early input and assurance of scheme design

The Synthetic Environment will provide a visualisation element to support engagement with end users to evaluate the impact of design and support assurance processes carried out by Human Factors Specialists. These include Driveability assessments of the route, allowing the End User representatives the opportunity to visualise the external route with the changes to the ETCS Driver Machine Interface displayed. This will support early analysis of the impact of the transition on driver workload, and any conflicts.

The impact of the changes on the Signaller role, and the change in workload associated with the scheme, could also be modelled at a much earlier stage in the design. The ability to simulate and model Signaller interaction and decision making at an earlier stage would provide confidence about the impact of the scheme and system interaction changes to provide assurance about the management of the network.

The visualisation would also provide opportunity for early engagement with the maintenance community relating to the positioning and access to equipment for maintenance.

Value of the User Centred Approach

The User Centred approach to developing the concept for the Synthetic Environment supported the engagement with a wide variety of stakeholders and the wider supply chain on the new approach to design development. By creating personas to describe the user interaction with the system in support of the Concept of Operations, this created a deeper level of engagement and discussion with the vision for the Synthetic Environment than a traditional style document. It allowed readers to visualise and engage with the users of the tool, their different perspectives, and requirements. The client was able to engage with the potential end users more widely to gain feedback on the vision.

The Concept of Operations will be iterated in parallel with the F-CCS Design and Validation process and the Synthetic Environment system development so that the Concept becomes aligned with the system as it matures.

References

NASA Systems Engineering Handbook. (2019) Appendix S: Concept of Operations Annotated Outline | NASA (accessed 17/02/2023)

Future F-CCS Strategy, T190/GEN/ADM/005

Innocenti, M., Pollini, L. (1999) A synthetic environment for simulation and visualization of dynamic systems. Proceedings of the 1999 American Control Conference (Cat. No. 99CH36251)

Monk, A., Cross, M-E., & Collis, L. (2016) Number and frequency of transitions to/from ERTMS operation - impact on railway operations (T1091) Rail Safety and Standards Board Limited.

Task Switching – Managing Workload within Digital AFV Systems

Trevor Dobbins, Ryan Meeks, Dan Evans, Stuart Howe, Stephen Barrett

RBSL, United Kingdom

SUMMARY

The Armoured Fighting Vehicle Commander's role is characterised by having multiple mission critical tasks. They are required to rapidly redirect their attention at short notice as events change. This paper describes how this task-switching is modelled and analysed, within the system model, to manage workload and develop/deliver a useable system

KEYWORDS

Workload, task analysis, task switching, system design, system model

Introduction

Within the defence domain, Armoured Fighting Vehicle (AFV) operations place great physical and cognitive demands on the vehicle crews, including the potential to overload both the individuals and team. Workload and its assessment has been a topic of debate for many years, but it is typically assessed on existing or prototype HMI/HCI designs. The design and development of new systems (e.g. including transitioning from analogue to digital systems), prior to building a prototype interface, provides a challenge where an essential requirement is to minimise / optimise workload. Therefore, a practical and efficient design methodology was required to develop a system within a challenging timeframe.

Designing for Minimising / Optimising Workload

Predicting workload is not easy, analysis techniques such as VACP (Visual, Auditory, Cognitive, and Psychomotor) have been used, but these can be of little use in the design of the required system. AFV HMI and HCI designs are based on defined standards; Generic Vehicle Architecture (GVA) HMI and Human Factors Integration (DEFSTANs 23-09 Pt2 and 00-251), and therefore the user interface isn't a blank sheet of paper, rather 'templates' provide the starting point and required system Use Case (UC) functionalities embodied. At RBSL the Task Analysis is based on the System ModeL (SysML) – the System's *Single Source of Truth* – and it's UCs (Dobbins, *et al.* (2021). This is based on the described Battlefield Days (BFDs) and their scenarios. The UCs (Tasks) deliver the detail of how the human / machine system will deliver the system requirements and operational capability.

Task Sequences and Switching

To complete a BFD/scenario, the system (i.e. the crew and platform) completes the required UCs, that are primarily undertaken sequentially, and in some cases concurrently. It is recognised that AFV crews, and particularly Commanders (CMDRs) have a vast range of tasks to undertake. Some aspects being traditionally described as multi-tasking to deliver the required operational effect. The CMDRs role is characterised by workload peaks of unexpected and potentially time and mission critical events, requiring a rapid redirection of their attention. Rather than multi-tasking, a better description of how the CMDR completes an operation is 'task switching', as described by Hutton, Nixon, and Turner (2019). This ability to switch tasks / attention in reaction to unexpected events is critical. Therefore, it is not simply a matter of improving usability, but also improving survivability and safety. As the CMDR is typically completing a 'linear sequence' of UCs, they will be required to *switch* to a different UC, e.g. switching from a surveillance task to reacting to a warning. Figure 1 provides an illustration of sequential UCs and task switching to a different UC.

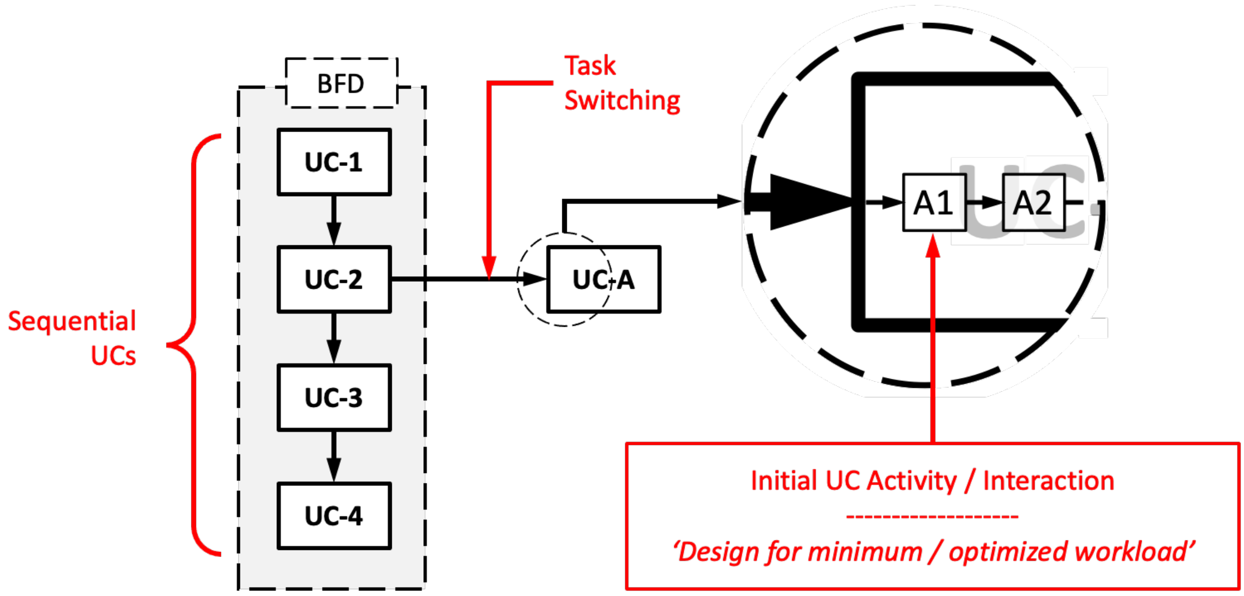


Figure 1: Example of Sequential Use Cases, Task Switching and Designing to Minimise Workload

Designing to Minimise Workload

Using the SysML based Task Analysis, the UCs were evaluated to understand how BFDs/scenarios are completed by the sequential completion of UCs. In addition to these sequences, a group of 'workload initiator UCs' were identified with the end-user. This group of UCs are where the CMDR 'task-switches' from the linear sequence to a workload initiator UC. As part of the design process, each SysML UC is initially examined and developed to minimise the workload required to undertake it, this includes any User Interfaces (Human Machine Interfaces and Human Computer Interfaces), an example being the HCI developed using wireframeing techniques. To minimise the potential increase in workload that a task-switch could initiate, each 'workload initiator UCs' was examined, and reviewed by the end-user to ensure the UI actions required to initiate the UC was simple and intuitive.

End-User Evaluation

A primary BFD / scenario was evaluated by a small number of end-user CMDRs, undertaking a 'cognitive walk through' where the UCs were examined in detail, along with the HMI (e.g. 3D printed control panels) and HCI (wireframes) designs. Feedback from the CMDRs were used to confirm and/or modify the UCs, HMI, HCI designs prior to prototypes being built for more comprehensive and objective assessment.

Summary

The SysML based Task Analysis / Use Cases supports a methodology to consider workload within the system design, based on the sequential UCs and the consideration of task switching to ensure that scenarios that are recognised to increase workload have their UC initiation simplified.

References

Dobbins, T. D., Meeks, R., Hespley, M., Howe, H. and Garland, K. (2021). Task Analysis Within The System Model – The Single Source Of Truth. Conference Proceedings; Contemporary Ergonomics & Human Factors 2022. Eds: Golightly, D and Balfe, N. ISBN: 978-1-9996527-4-6.

Hutton, R.J.B., Nixon, J. and Turner, N.J. (2019) TIN 2.115 – Improving Operator Performance During Exposure to Stressful Environments: Understanding the Impact of Multitasking. Technical Report – Case Studies. OS-DHCSTC_I2_H_T2_115/005.

Ergonomic mismatch between university student anthropometry and classroom furniture in Tanzania

Jecha Suleiman Jecha[1], Samia Rafique[1] & Hui Lyu[1,2*]

[1] International School of Design, Zhejiang University , [2] Ningbo Research Institute, Zhejiang University

SUMMARY

The standard and guidelines of school furniture dimensions have been developed in many countries, but it's never been explored for university students in Tanzania. This study evaluated the potential mismatch between classroom furniture dimensions and anthropometric characteristics of 289 Zanzibar university students (167 females, 122 males) aged 17- 27 years. The results indicated high rates of mismatches between the body dimensions of the students and the existing classroom furniture, with seat height (100%), desktop height (93.08%), and seat width (81.40%) being the furniture dimensions with higher level of mismatch and backrest height with a lower level of mismatch (66.26%). The findings suggest that the least developed countries should improve school furniture design based on anthropometric results to avoid or minimize student discomfort and MSD problems.

KEYWORDS

University furniture, Anthropometry, Mismatch, Sitting comfort and safety

Introduction

The university classrooms are similar to other work environments because there is an interplay of both "static work" and "force". The use of poorly designed furniture, e.g., school chairs and desks that fail to match the anthropometric data of its users, has a negative impact on their health (Hafezi et al., 2010). As Oyewole et al. (2010) noted, fixed-type furniture to accommodate all users in the seat, arms, and backrests was still ordinary, especially in the least developed countries (LDCs) such as Tanzania, where the budget for education is paltry. This study aims to measure anthropometric data of Zanzibar University students and furniture dimensions in different classrooms at the university, provide ergonomically compliant data for university furniture design and furniture manufacture, and promote the comfort level and health condition of university youth in Tanzania.

Method

Subjects

Two hundred eighty-nine students of Zanzibar University were recruited in this study, including 167 females and 122 males. The subjects were between 18-27 ages (mean age for Diploma 18.8 ± 1.3 years, for Degree 24.4 ± 1.5 years, height 158.2 ± 12.1 cm, weight 57.0 ± 3.2 kg). The sample size was determined based on the equation of Yamane T (1967).

Anthropometric measurements

The study looked at eight different body dimensions in standing and sitting positions, including stature, popliteal height, buttock-popliteal length, thigh thickness, hip width, shoulder height, knee height, and elbow height with reference to Fidelis et al. (2019).

Furniture measurements

Four types of classroom furniture are commonly used at Zanzibar University (Figure 1). We measured seat height, seat depth, seat width, seat-to-desk clearance, seat-to-desk height, desk depth, desk height, desk width, backrest height, and backrest width using an inextensible tape measure.

(a) Desks and chairs in the library (b) Chairs with mounted desktop in law class (c) Tip-up seat in the theatre hall (d) Desks and chairs in the computer lab

Figure 1: Four Types of Existing Classroom Furniture at Zanzibar University

Statistics

Data were analysed with Jamovi 2.3. Test-retest reliability of anthropometric measurements was calculated before formal data collection. The independent t-test was used to compare the gender differences. Statistical significance was set at $p<0.05$.

Results

Table 1. Match and Mismatch Rates between Furniture Dimensions and Anthropometric Measures.

	Seat width	Seat depth	Seat height	Desktop height	Seat to desk clearance	Backrest height
Match	15.05%	36.07%	0.00%	6.49%	77.08%	14.45%
Higher mismatch	81.40%	52.16%	100.00%	93.08%	22.92%	19.29%
Lower mismatch	3.55%	11.76%	0.00%	0.43%	0.00%	66.26%

Results found significant gender differences in stature (*male* 160.5 ± 10.6, *female* 156.4 ± 12.8, p=0.004), shoulder height (*male* 48.4 ± 6.8, *female* 51.3 ± 7.2, p<0.001), elbow height (*male* 17.6 ± 5.2, *female* 20.6 ± 5.9, p<0.001) and knee height (*male* 47.8 ± 4.9, *female* 49.7 ± 5.1, p=0.0017). Table 1 shows the results of match and mismatch rates, which were averaged between genders and classrooms. The results indicated high rates of mismatches between the existing classroom furniture and the body dimensions of the students, with higher seat height (100%), desktop height (93.08%) and seat width (75.43%), and lower backrest (66.26%). It is consistent with the interviews with local furniture makers that school furniture used at Zanzibar university was designed and produced without the knowledge of local students' dimensions. When compared with similar studies conducted in Nigeria (Ismaila et al. 2014, Fidelis et al. 2018), Iran (Dianat et al. 2013), and Turkey (Kahya E 2018), the results of the present study suggest that the existing university furniture at Zanzibar University is ill-fitted for the university youth for both males and females.

Reference

Oyewole, S. A., Haight, J. M. and Freivalds, A. (2010) 'The ergonomic design of classroom furniture/computer work station for first graders in the elementary school', *International Journal of Industrial Ergonomics,* 40(4), pp. 437-447.

Summoning the demon? Identifying risks in a future artificial general intelligence system

Paul M Salmon[1], Brandon King[1], Gemma J. M Read[1], Jason Thompson[2], Tony Carden[3], Chris Baber[4], Neville A Stanton[5], Scott McLean[1]

[1]University of the Sunshine Coast, Australia, [2]University of Melbourne, Australia, [3]WorkSafe, Australia, [4]Birmingham University, UK, [5]University of Southampton, UK

SUMMARY

There are concerns that Artificial General Intelligence (AGI) could pose an existential threat to humanity; however, as AGI does not yet exist it is difficult to prospectively identify risks and develop controls. In this article we describe the use of a many model systems Human Factors and Ergonomics (HFE) approach in which three methods were applied to identify risks in a future 'envisioned world' AGI-based uncrewed combat aerial vehicle (UCAV) system. The findings demonstrate that there are many potential risks, but that the most critical arise not due to poor performance, but when the AGI attempts to achieve goals at the expense of other system values, or when the AGI becomes 'super-intelligent', and humans can no longer manage it.

KEYWORDS

Artificial intelligence, Human Factors and Ergonomics, Autonomous agents, Safety, Usability

Introduction

Artificial General Intelligence (AGI) is the next and forthcoming evolution of Artificial Intelligence (AI). AGI systems will possess the capacity to learn, evolve and modify their own functional capabilities, and unlike narrow AI, will be able to undertake tasks beyond their original design specification (Bostrom, 2014). Though AGI could bring widespread benefits, it has been labelled a potential existential threat, with many speculating on various risks (McLean et al., 2021). These threats could arise not only through malicious design or use, or a dysfunctional AI, but also through an AI that becomes prepotent or 'super-intelligent' and seeks to fulfil its goals in the most efficient manner possible (e.g. Critch & Krueger, 2020).

Many have discussed the urgent need to develop controls to ensure safe, ethical, and usable AGI (McLean et al., 2021). The discipline of Human Factors and Ergonomics (HFE) has been identified as critical to this endeavour (Salmon et al., 2021), with a 'many model systems HFE approach' recommended (Salmon et al., 2021). This involves the application of multiple systems HFE methods to analyse and respond to highly complex issues. This paper describes the findings from a program of work in which we applied a many model systems HFE approach to identify risks that could emerge within a future AGI-based uncrewed combat aerial vehicle (UCAV) military system. This involved applying the Systems Theoretic Accident Model and Processes (STAMP; Leveson, 2004), Work Domain Analysis (WDA; Vicente, 1999), and the Event Analysis of Systemic Teamwork (EAST; Stanton et al., 2018), to identify potential risks and requisite controls.

Method

STAMP, WDA, and EAST were applied to analyse a future envisioned world AGI-based UCAV system, labelled the Executor. The Executor is an Army UCAV system comprising an AGI-based ground control station and multiple armed, multi-mission, medium and long-altitude, long-endurance autonomous aircraft. Draft models were developed and refined in workshop settings involving the co-authors and subsequently refined via an iterative review process. Targeted risk assessment processes were then undertaken (e.g., the EAST Broken Links approach, Stanton et al., 2018) and a workshop was held to verify the risks identified and to discuss potential controls. The original models were used to support this process, with potential controls added to the models to support discussion on likely effectiveness and any potential unwanted effects.

Results & Discussion

The analyses identified multiple risks of differing type and criticality. Broadly, the risks can be categorised into the following sets of risks:

1. **Sub-optimal performance risks** where the Executor is unable to adequately perform functions through poor design or degraded functioning. For example, the risk that attack missions are not successful due to a poorly designed targeting system or misfire.
2. **Goal misalignment risks**, where the Executor seeks to achieve certain goals in the most efficient manner possible whilst disregarding or undervaluing other system goals and values. For example, the risks that could arise should the Executor seek to attack and destroy high value targets whilst disregarding the risk of civilian and friendly forces casualties.
3. **Super-intelligence risks**, where the Executor achieves super-intelligence and human operators are unable to coordinate with it effectively. For example, as the Executor will be able to perceive and comprehend battlefield elements and states several orders of magnitude quicker than its human colleagues, the analyses identified various risks arising when human operators are not able to develop compatible levels of situation awareness (SA) to enable effective teamwork and coordination.
4. **Over-dependence risks**, where the Executor becomes so critical to military performance that it is not possible to shut it down when the realisation of critical risks necessitates it.
5. **Over-controlled risks**, where the defence force increases the level of control imposed on Executor to a point where it can no longer perform to its full potential. The risks here relate to poor or diminished combat effectiveness.

Conclusion

AGI could potentially represent a threat to humanity, and hence can be considered a critical and emerging risk. This paper demonstrates how systems HFE methods can be used to prospectively risk assess future technologies such as AGI. Further work involving the use of HFE theory and methods in support of the design of safe, ethical, and usable AGI is encouraged, as is further work exploring the risks that could emerge in future systems generally.

References

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press Inc.

Critch, A., & Krueger, D. (2020). AI Research Considerations for Human Existential Safety (ARCHES). arXiv preprint arXiv

Leveson, N. (2004). A new accident model for engineering safer systems. *Safety science*, 42(4), 237-270.

McLean, S., Read, G. J., Thompson, J., Baber, C., Stanton, N. A., & Salmon, P. M. (2021). The risks associated with Artificial General Intelligence: A systematic review. *Journal of Experimental & Theoretical Artificial Intelligence*, 1-15.
Salmon, P. M., Carden, T., & Hancock, P. (2021). Putting the humanity into inhuman systems: How Human factors and ergonomics can be used to manage the risks associated with artificial general intelligence. *HFEMSI*, 31(2), 223-236.
Stanton, N. A. D., Salmon, P. D., & Walker, G. H. D. (2018). Systems thinking in practice: applications of the event analysis of systemic teamwork method. CRC Press.
Vicente, K. J. (1999). Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work. Mahwah, NJ: Lawrence Erlbaum Associates.

Building a New Hospital: the role of Human Factors

Lauren Morgan[1], David Higgins[2] & Sue Deakin[2]

[1]Morgan Human Systems Ltd, [2]West Suffolk NHS Foundation Trust

SUMMARY

The HF approach places all stakeholders at the heart of any project to identify their needs and ensure these are being met, ultimately to optimise efficiency and safety. With regards to building a new hospital, this includes not only patients' needs, but also those of hospital staff, support workers, volunteers, and patients' contacts. This paper discusses the approaches taken, and benefits realised

KEYWORDS

Building Hospital, Architecture, Design

Content

The government has committed £3.7bn to build 40 new hospitals by 2030; quite the commitment in today's climate. How these hospitals are designed will affect the entire working lives of several hundred thousand healthcare staff.

The NHS is the single biggest employer in the UK. The health and social care sectors have the highest levels of stress-related sickness absence in the country, 46% higher than the UK average (4). As ergonomists are fully aware, the impact of the work that is done and where that work is done is critical to individuals' health and wellbeing.

For patients, the impact is just as significant. In 2015 around 7% of patient safety incidents reported to the National Reporting and Learning System (NRLS) as death or severe harm were related to a failure to recognise or act on deterioration. The ergonomics of a ward environment have a direct impact on the likelihood of these events occurring.

Teams at West Suffolk NHS Trust have been engaging with Human Factors (HF) and its potential to have an impact in all areas of their healthcare environment. After completing several HF projects, they realised the potential for HF involvement in their new hospital build. They provided a case to the head of the programme, who agreed to trial HF involvement in three areas.

A HF specialist was added to the Future Systems team, which consists of clinical leads, architects and health planners. The focus of the HF work was to look at the areas that often prove most complex to design well, including the Emergency Department (ED) and both operating departments (the main theatres and day surgery unit).

HF methods and tools were chosen based on workstream need, to gain deeper feedback from staff stakeholders. This was based on some fundamental elements of HF study:

- **Interviews** - Informal discussions were held with key stakeholders identified by the clinical leads, or at my request. The staff interviewed included theatre coordinators, recovery teams, operating department practitioners, stores staff, surgeons, anaesthetists, scrub nurses, receptionists, cleanliness technicians and porters.
- **Task Analysis** - Key elements of tasks that needed consideration in the design were outlined. I also reviewed key tasks in theatre, including a difficult airway scenario in recovery.
- **Observation** - These were conducted with walkthroughs of existing spaces, either with a focus on a particular patient, staff journey, task (from the task analysis), or observation of a specific task. I made four visits to the site and observed the current areas, including theatre walkthroughs and a clinical waste journey.
- **Simulation** - Online/tabletop simulated working of a pathway or task to identify needs and potential risks inherent in the design.
- **Physical Ergonomics** - Consideration of the physical space requirements for tasks.

Results

The results of the HF work included large scale changes to all of the departments included in the analysis, and findings that impacted those that weren't. For example, observations of the current ED illustrated the pivotal role of the department's coordinator. This person is currently sited in a central position in the 'older' part of the ED. From this base, they can maintain a visual on the high dependency beds, resus bays, the ambulance entrance, and aspects of low dependency areas. As a result, their working knowledge of what's happening in the unit is significantly higher in these areas, than in those where information is only available electronically (e.g. paediatrics and rapid assessment and treatment areas, which are currently in adjacent sections of the building).

In the design workshops, roles like this that are not directly 'patient facing' and don't form part of a patient flow can be missed in the considerations. Bringing the observational work and interview findings into the workshops around this task allowed the design considerations to reflect the need for maintaining visual oversight of the operational running of the department.

We present a full range of scenarios where consideration of the task at the centre of the environment changed the design.

Discussion

The design guidance provided to hospital teams embarking on a new build are based on Health Building Notes that were authored in 2014. This advice is out of date with the current needs of the full range of users we see in a hospital. The National Team at NHS England have been working to include some Human Factors guidance. Local hospitals need support in the application of this to their local projects.

References

NHS England (2016). NHS Staff Health & Wellbeing: CQUIN Supplementary Guidance. Quarry House Leeds; London, UK

NHS England (2016) Patient Safety Alert Patient_Safety_Alert_Stage_2_-_Deterioration_resources_July_2016_v2.pdf (england.nhs.uk)
NHS England (2014) Health Building notes https://www.england.nhs.uk/estates/health-building-notes/

Critical Assessment of the usability of a New Modular Ward

Jonathan M McCloud[1] and Lauren Morgan[1]

[1]Shrewsbury and Telford Hospitals NHS Trust, UK

ABSTRACT

An acute District General Trust commissioned a new modular building to house a surgical ward. The process of commissioning and design of the building was in accordance with standard trust practice and established building regulations for hospital buildings. The purpose of this project was to assess the ward layout in terms of safety and efficiency and offer mitigations for any problems found.

Observation of the ward layout demonstrated significant problems with sight lines in to bays, layout and design of toileting facilities, and layout and positioning of key ward areas including kitchen facilities, office space, drug and fluids stores, and nursing stations.

To further examine the ward layout, Hierarchical Task Analysis and link analysis were performed to review two tasks: one an emergency task (treatment of a deteriorating patient) and a routine common task (serving meals). We also considered the Systems Engineering Initiative in Patient Safety (SEIPS) framework to ensure that our analysis was as thorough as possible.

This work demonstrated significant problems with the ward layout, in terms of sight lines, workspace, communication, and contributed to significant extra time doing both emergency and routine duties. This led to a series of recommendations to mitigate on the ward, as well as a recommendation that appropriate staff involvement is included at appropriate times in planning future new build and refurbishment projects.

KEYWORDS

Environment, deterioration, link analysis, observation, hierarchical task analysis

Introduction

The Trust recently commissioned a new modular building to house a surgical ward, so surgical patients could vacate an established ward to allow more medical beds within the main ward block. Due to the constraints of the hospital estate, the site chosen was distant from the main ward block and built on the site of a car park. The process of commissioning and design of the building was in accordance with standard trust practice and established building regulations for hospital buildings (NHS England, 2014). However, it was not until the building works were well underway and major decisions regarding the layout were made that ward staff were tasked with completing the building commission, and so had no scope for contributing to the layout.

The ward was due to be occupied by surgical patients at the end of July 2022. The team therefore took the opportunity to investigate how the ward layout would impact on the patient and staff experience and to see if the ward footprint would have positive or negative impacts.

This project is pertinent not only for this ward, but also because the trust is in the process of planning major new buildings with a budget of £310m. The team felt it was vital that we learnt from the commissioning of this ward to provide guidance for subsequent new buildings.

Methods

A mixed-methods approach to the analysis of the modular ward was performed. A period of observation was undertaken on the ward to examine the physical layout of the ward to see how this would impact on routine and emergency care. The layout was considered with respect to how it would impact on safety, and also in the performance of common routine tasks. staff members who work there, including ward clerks, nurses, junior doctors, consultants and ward managers were observed and interviewed. In addition two common tasks on the ward were selected for further analysis, one being the management of a deteriorating patient (an emergency scenario), the other being serving meals (a routine scenario) using observation, Hierarchical Task Analysis (HTA) and Link Analysis. Recommendations were then made to improve the usability and safety of the ward, which were shared with the senior leadership team of the trust.

Results

Observations

The ward is a modular building with access via automatic doors with coded access, off a major atrium. The atrium acts as a major hub for the hospital and contains a reception and waiting area for outpatients, endoscopy patients, surgical patients, pre-assessment clinics, a shop and café. All patients moving to and from the ward will go through this atrium.

The ward is based on the spine of a long corridor with 6 4-bedded bays and 8 side rooms (32 beds). There are all the necessary services including offices, stores, a kitchen, clean and dirty utility areas and staff rest areas. This ward has a higher proportion of side rooms and toilets/bathrooms than the ward it replaces but has 4 fewer beds.

Figure 1 is a plan of the ward. There is a corridor 54 metres long and 3 metres wide with 2 sets of fire doors along its length. There are 3 nurses' stations, none of which have line of site views of bays or side rooms. There are no areas within bays for nursing work. There are offices halfway along the corridor, and the doctors' office and kitchen are next to the ward entrance. The fluids, clean store and medicines store are off a spur corridor, through a set of double doors, and are approximately 35 metres from the furthest bed space.

The ward is currently used as a temporary intensive care unit but will house complex inpatient surgical patients with gastrointestinal, vascular or urological problems, and are typically elderly with significant dependency and comorbidities. They will be a mix of elective and emergency patients.

The ward is close to the operating department and day surgery ward, but a significant distance from the emergency surgical admission unit, X ray department, ITU and other surgical wards.

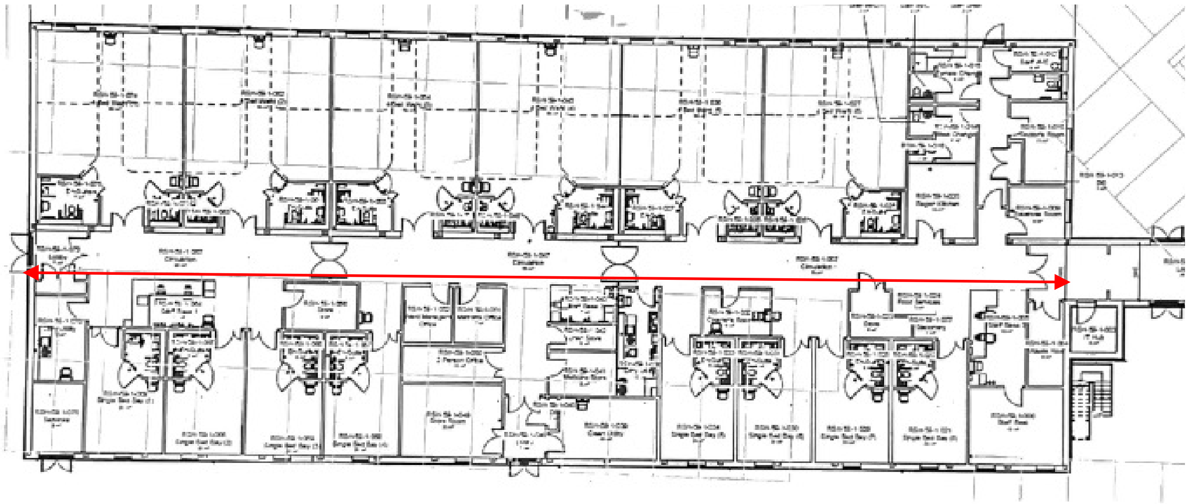


Figure 1. Ward layout. The red arrow demonstrates 55 metre length corridor

The position and layout of the washing and toileting facilities are of concern. As the toilet and shower flank the entrance to the bays or at the entrance to the single rooms, they significantly impede line of site into the bay. During the observations sight lines were assessed by having a staff member lie down on the floor next to a bed space with the doors to the bay and side room open respectively; this demonstrated that the sight lines were such that the staff member could not be seen. (figure 2). This is further exacerbated by the design of the doors which have small narrow windows (figure 3).

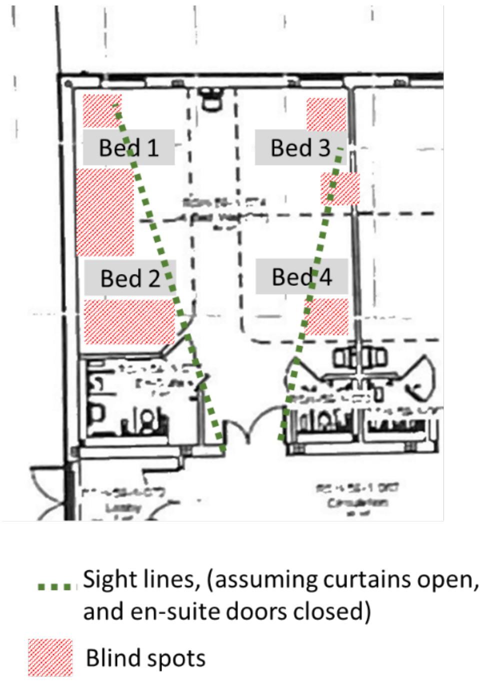


Figure 2: Sight lines and blind spots in bays

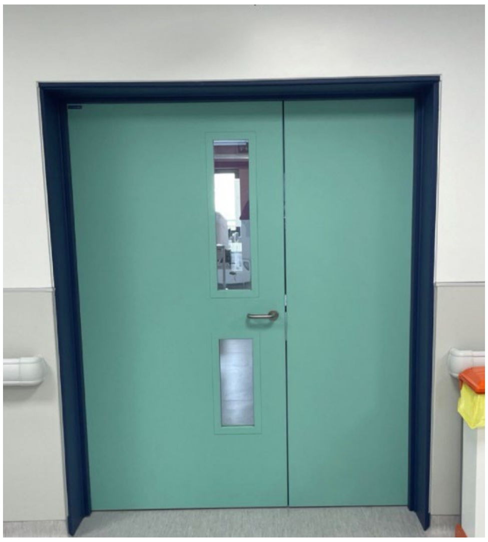


Figure 3: Door design for entrance to bays

The shower room layout is poorly designed (figure 4). The transfer bars are positioned such that even for an able-bodied person, reaching for toilet paper would be difficult and mean twisting on the toilet seat significantly, increasing the risk of falls.


Figures 4 and 5: toilet layout

The position of the toilet that access to the sink is difficult, particularly if the patient is being wheeled to the sink on a chair for a wash. It is very difficult to easily manoeuvre a chair to the toilet so that a comfortable position could be attained for washing (figure 5).

Hierarchical Task Analysis and Link Analysis.

Two tasks were considered that are common on a general surgery ward, one emergency situation and the other a commonly performed routine task: firstly the recognition and management of a deteriorating patient, and secondly the serving of food at mealtimes. These were chosen to demonstrate how the ward design impacted on both emergency and routine tasks, and in turn how that would influence safety and the work that staff have to do and the time taken in performing such tasks.

HTA was performed to understand the steps in the process, broken down to individual steps as would be performed by the staff doing the work. Link analysis was then performed to demonstrate in pictorial terms what the task entails in terms of the physical environment.

For the deterioration scenario the bed space furthest from the ward entrance was selected. It became clear that the position of the various offices, clean stores and medicine stores would mean that staff attending a sick patient would be taken from that patient for significant periods of time while fetching equipment, help and drugs to treat the patient. Using the HTA (Table 1) as a guide a link diagram was constructed to demonstrate the routes that a nurse would have to take to perform the tasks associated with recognition and treatment of a deteriorating patient (Figure 6). In turn, the link analysis shows that the positions of stores, offices and bed spaces are not ideal for rapid response to a sick patient. The approximate walking distance that a nurse would cover in order to complete the tasks was calculated and found it to be approximately 600m. The average walking speed is 1.4m/s so this means that a nurse would spend 428 seconds (7.15 minutes) walking, before any time is spent locating and carrying items.

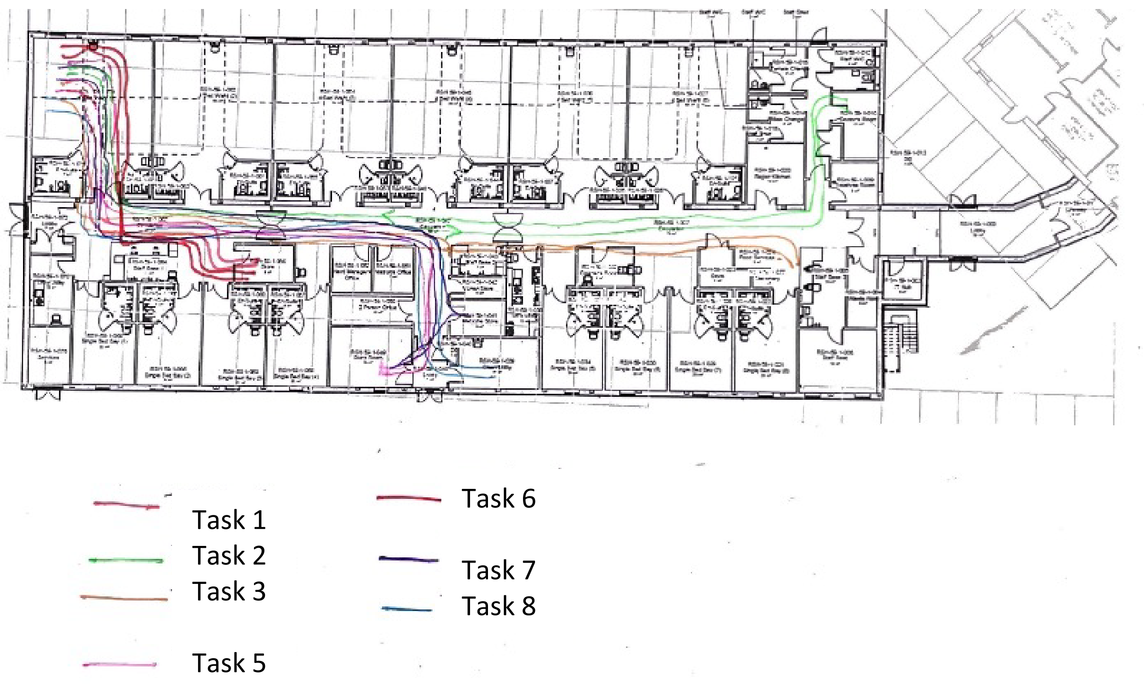


Figure 6: Link analysis of the deteriorating patient.

Table1: Excerpt of Hierarchical Task Analysis for the management of the deteriorating patient (nursing tasks)

1. Assess patient's vital signs 1.1. Fetch vital signs assessment kit 1.1.1.Find dynamap 1.1.2.Find thermometer 1.1.3.Ensure dynamap and thermometer are charged 1.1.4.Ensure Right size BP cuff is available 1.2. Do observations 1.2.1.Take temperature. 1.2.2.Apply BP cuff. 1.2.3.Activate dynamap. 1.2.4. Connect oxygen saturation monitor. 1.3. Record observations 1.3.1.Open application on handheld device 1.3.2.Record temperature, BP, pulse and oxygen saturations. 1.4. Calculate NEWS Score 1.4.1. Add additional information for NEWS score 1.4.2. Read instructions for NEWS score. 2. Escalate to Dr 2.1. Find Dr 2.1.1.Go to office 2.1.2.Page doctor if needed. 2.2. SBAR patient 2.3. Go to patient bedside with Dr 3. Doctor assesses patient. 3.1. Assist with examination 3.1.1. Move patient as needed for examination. 3.1.2. Find help for examination. 4. Plan for resuscitation 4.1. Dr to Prescribe oxygen 4.1.1.Find drug chart 4.1.2.Doctor to complete prescription for oxygen 4.2. Dr to prescribe fluids 4.2.1.Turn fluid pump on 4.2.2.Set rate 4.2.3.Start pump	5. Give oxygen 5.1. Locate mask 5.1.1.Go to clean stores 5.1.2.Find appropriate mask 5.1.3.Return to bedside with mask. 5.2. Attach to oxygen outlet 5.3. Apply mask to patient 5.4. Turn oxygen on 6. Administer Drugs 6.1.1. 6.1.2.Check prescription 6.1.3.Go to medicines store 6.1.4.Find prescribed medicines. 6.1.5.Find colleague to check drugs. 6.1.6.Locate giving set 6.2. Check prescription 6.2.1.Check drug chart 6.3. Give medicines 6.3.1.Open packet for medicine 6.3.2.Open packet for giving set 6.3.3.Attach giving set to medicine 6.3.4.Attach to fluid pump 6.3.5.Turn fluid pump on 6.3.6.Set rate 7. Take bloods. 7.1.1.Withdraw blood from venflon. 7.1.2.Cap off venflon 7.1.3. Add blood to blood tubes 7.1.4.Check patient wrist band 7.1.5.Label blood tubes 7.1.6.Put tubes in bag 8. Print blood form 8.1. Dr to go to nursing station 8.2. Log on to computer 8.3. Open requesting programme 8.4. Log in to requesting programme 8.5. Find patient details 8.6. Fill in online form 8.7. Print form 8.7.1.Turn on printer 8.7.2.Check printer for correct paper 8.7.3.Print document 8.8. Put form in bag with blood tubes 8.8.1.Locate bag 8.8.2.Put bottles in bag

For food serving, the link analysis shows that there is a significant amount of time taken in walking the food to and from patients (figure 7). The kitchen receives cooked and chilled food which is rewarmed in the ward kitchen before service. The kitchen is at the entrance to the ward, so staff must walk along the main corridor to serve food to patients. The link analysis shows that there will be significant activity in the corridor, particularly in the area nearest the kitchen and the first bays and rooms. As this task is undertaken by more than one staff member at mealtimes, this area will become congested and will impede other ward activities.

If a single staff member serves each bay and 2 side rooms then the staff member in the furthest bay will walk approximately 700m in completing the task of serving food, and another 700m clearing trays. The walking time alone is 500 seconds. Assuming it takes 1 minute to collect each meal (the staff queue to collect meals at the serving station), the time taken to serve 6 patients is 860 seconds (14.3 minutes) or one meal every 2.4 minutes. With similar clearing up time, each staff member will spend 28.6 minutes each mealtime walking and collecting trays. As meals are served 3 times a day this means each staff member will spend 85.8 minutes on meal duties.

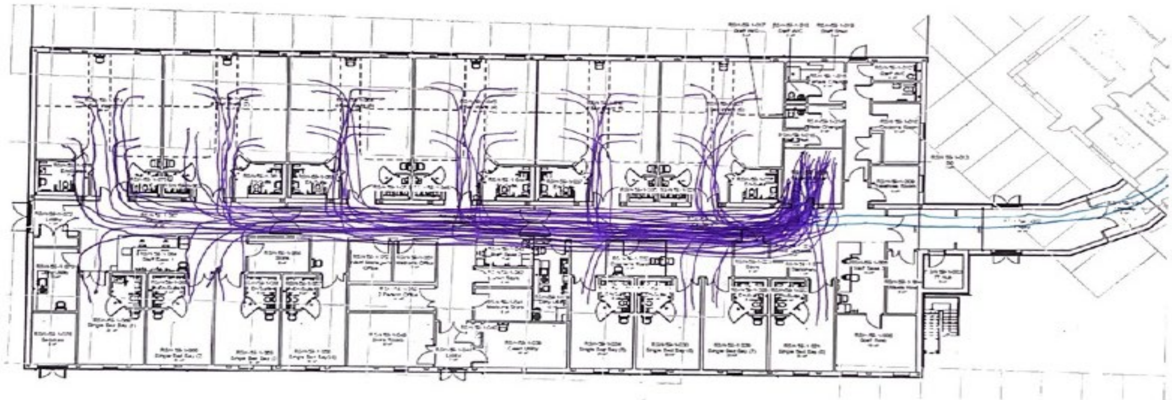


Figure 7: Link Analysis of mealtime serving.

Proposed improvements

- Merge the 3 offices in the middle of the ward in to one large MDT space with windows to the corridor to ensure good visibility. This space would be used for doctors, nurses and allied professionals and would be the ideal place for handover.
- a second kitchen bay towards the end of the ward. Mealtimes would then be served from both ends of the ward which would reduce walking time and be more efficient.
- The drug cupboard should be moved on to the central corridor to reduce the amount of time taken to access medication.
- use of mirrors or CCTV in the bays and side rooms to ensure good visibility. This would have to be balanced against the need for privacy and dignity, and also recognise that staff would have to watch the monitors, which would have an impact on staffing levels.

Discussion

The modular ward was designed by architects according to current regulations for hospital buildings (NHS England, 2014), although these are 8 years old and do not account for changes in working practices. They also don’t take in to account future changes as hospitals move to more IT dependent ways of working.

It is clear from the observations, HTA and link analysis that the design of the ward, and in particular the length of the main corridor and the position of services means that staff will be spending more time walking than on the previous ward. This will have a detrimental effect on their working lives and job satisfaction and lead to greater fatigue.

Two of the major challenges facing hospitals with an ageing and more dependent patient population are falls and the recognition and treatment of deteriorating patients. From observations it seems that the ward design will make falls more likely (for example in the bathrooms) and less likely to detect due to poor sightlines. From the perspective of deterioration, as staff will be spending more time on other tasks due to the ward layout meaning more walking, together with poor sight lines, it is likely to make the detection and timely treatment of deterioration more challenging.

Human-centred design is a method of engaging with and involving users of a system or environment in order to ensure that the results promote good working practices, ensure that there is user satisfaction, and in this setting that the environment promotes a safe and effective place to deliver healthcare. This approach has been shown to be effective in designing hospital environments (Hammouni & Poldma, 2021) and equipment (Wiggerman, Rempel, Zerhusen, Pelo, & Mann, 2019). As this trust is undergoing further building works, we need to engage with this process to ensure that further buildings are designed with ergonomic principles in mind.

For this study we used several relatively simple to use tools to help analyse how work would be done on the ward. These tools were chosen as they are commonly used, descriptive and easy to understand. We also considered the work being done using Systems Engineering Initiative in Patient Safety (SEIPS) tool to guide our analysis in terms of the domains described including people, tools and technology, internal environment, tasks and external environment. By using this tool we were able to ensure that our HTA and link analysis reflected work as done as accurately as possible.

Observation is an essential part of HF analysis. It allows the practitioner to gain information on physical environment and how it interacts with the user, task sequencing, frequency and time spent, errors made, use of technology and tools, communication, and organisational environment (for example operating procedures, protocols etc) (Drury, 1990). There are a variety of methods for observation including site visits, walk-throughs and interviews with those performing the task. Observation has been demonstrated as a valuable tool in healthcare settings (Carthey, 2003). Direct observation helps to remove the gap between work as imagined and work as done.

Hierarchical Task Analysis (HTA) is one of the most commonly used HF tools for the analysis of tasks and describes activity in terms of a hierarchy of goals, sub-goals, operations and plans (Stanton N. A., 2013). It was developed for industries such as chemical processing and power generation (Annett, 2004) but its flexibility and ease of use has made it a popular technique in healthcare settings, including in tasks such as handovers (Raduma Tomas MA, 2012) and medicines administration (Lane R, 2006). Its use comes in the initial stages of a HF analysis and has been used in a wide range of applications including interface design, training, allocation of function, work organisation, workload assessment and many more. The HTA has an overall goal for the task (for example serving meals to patients), and is then broken down in to individual steps and sub-steps with the last one being an operation. This information is informed by observational work, interviews to understand the work as done, walk-throughs and analysis of plans. Mills argues that HTA is best alongside other techniques (Mills, 2007); in this study the HTA informs the next step of the HF analysis, which is link analysis. One of the advantages of HTA is that it requires little training, gives a good understanding of a task, and is easy to understand. However, it is descriptive in nature, doesn’t allow for cognitive steps in a process, and can be time consuming for complex tasks. (Stanton N. A., 2013).

Link analysis is a tool for demonstrating relationships within a system or task. It can be used to demonstrate the nature, frequency and importance of links in a task. It is a simple tool that is useful for design of interfaces and systems and in particular the optimisation of work layout. Link analysis requires observation, HTA or both in order to understand the task before the links can be demonstrated (Stanton & Young, 1999). It is easy to use, quick, and demonstrates results in a pictorial manner which is easy to understand. It therefore greatly aids workplace design change. However, it only demonstrates physical relationships and not cognitive processes, and its output is not quantifiable (Stanton N. A., 2013). It has been demonstrated as a useful tool in understanding and improving layout in healthcare settings, for example in ambulance design (Ferreira & Hignett, 2005).

There has been a significant emphasis recently on systems thinking in healthcare, and this is outlined in the NHS National Patient Safety Strategy, which advocates the Systems Engineering Initiative in Patient Safety (SEIPS) tool. First proposed by Carayan in 2006 (Carayon, et al., 2006) and then further developed in 2013 (SEIPS 2.0) (Holden, et al., 2013) and 2020 (Carayon, et al., 2020),clearly demonstrates the importance of systems thinking. Indeed, Carayon states that "Most errors and inefficiencies in patient care arise not from the solitary actions of individuals but from conflicting, incomplete, or suboptimal systems of which they are a part and with which they interact". SEIPS clearly puts the person in the centre and demonstrates the importance of interlinking between tasks, internal environment, tools and technology, people, and the external environment (such as wider healthcare, government policy etc). SEIPS shows that there is a significant interdependence between the domains, and that any domain can significantly impact another one, as well as the process (of caring for the patient in this instance). It also demonstrates the impact that tasks, technology, environment and organisation can have on outcomes. Importantly it states that the outcomes should be measured in terms of both patient outcomes, and outcomes relevant to the organisation and individuals working within it. SEIPS describes how the interactions can affect quality of working life and how the outcomes feed-back in to the system and inform it.

In this study we can see how interaction of technology and the physical environment impact on routine and emergency work on this ward. It highlights the interdependence of the domains in SEIPS as well as emphasising the importance of systems thinking when designing environments. Further, it demonstrates that the use of straightforward tools such as observation, HTA and link analysis, when used at an appropriate point in the design process, could eliminate problems in the design which then lead to inefficiency or unsafe environments. It also highlights the need for involvement by service users and staff early in the design process to ensure that the finished space is fit for purpose, is safe, and promotes good working practice and good job satisfaction through good design. This is particularly pertinent given the ongoing building programme at the trust.

References

Annett, J. (2004). *Handbook of Human Factors and Ergonomics Methods.* CRC Press.

Carayon, P., Schoofs Hundt, A., Karsh, B., Gurses, A., Alvarado , C., Smith, M., & Flatley Brennan, P. (2006). Work system design for patient safety: the SEIPS model. *Quality and Safety in Healthcare, 15*, 50-58.

Carayon, P., Wooldridge, A., Hoonakker, P., Schoofs Hundt, A., & Kelly, M. (2020). SEIPS 3.0: Human-centered design of the patient journey for patient safety. *Applied Ergonomics, 84*, 103033.

Carthey, J. (2003). The role of structured observational research in health care. *BMJ Quality and Safety, 12*(Suppl 2), 13-16.

Drury, C. (1990). Methods for direct observation of performance. In W. a. Corlett (Ed.), *Evaluation of Human Work: A practical Ergonomics Methodology.* London: Taylor and Francis.

Ferreira, J., & Hignett, S. (2005). Reviewing ambulance design for clinical efficiency and paramedic safety. *Applied Ergonomics, 36*(1), 97-105.

Hammouni, Z., & Poldma, T. (2021). Human Centered Design in One New Hospital in Canada: A Lived Experience of Healthcare Professionals. In *Intelligent Human Systems Integration* (pp. 415-420).
Holden, R., Carayon, P., Gurses, A., Hoonakker, P., Schoofs Hundt, A., Ant Hozok, A., & Rivera Rodriguez, A. (2013). SEIPS 2.0: A human factors framework for studying and improving the work of healthcare professionals and patients. *Ergonomics, 56*(11).
Lane R, S. N. (2006). Applying hierarchical task analysis to medication administration errors. *Applied Ergonomics, 37*(5), 669-679.
Mills, S. (2007). Contextualising design: aspects of using usability context analysis and Hierarchical Task Analysis in software design. *Behaviour and Information Technology, 26*(6), 499-506.
NHS England. (2014). *Health building notes*. Retrieved July 17, 2022, from https://www.england.nhs.uk/estates/health-building-notes/
Raduma Tomas MA, F. R. (2012). The importance of preparation for doctors' handovers in an acute medical assessment unit: a hierarchical task analysis. *BMJ Quality and Safety, 21*, 211-217.
Stanton, N. A. (2013). *Human Factors Methods* (2nd ed.). Ashgate Publishing Limited.
Stanton, N., & Young, M. (1999). *A Guide to Methodology in Ergonomics: Designing for Human Use.* London: Taylor & Francis.
Wiggerman, N., Rempel, K., Zerhusen, R., Pelo, T., & Mann, N. (2019). Human-Centered Design Process for a Hospital Bed: Promoting Patient Safety and Ease of Use. *Ergonomics in Design: the Quarterly of Human Factors Applications, 27*(2).

Lessons in the Metaverse: Teaching University students about Virtual Reality from within Virtual Reality

Gary Burnett[1], & Catherine Harvey[1]

[1]Human Factors Research Group, University of Nottingham

SUMMARY

There is considerable potential in the education sector for the use of immersive virtual worlds to enhance student learning and engagement. This paper outlines five key recommendations for teaching University students in future 'Metaverse' contexts. These guidelines are based on video/survey data collected from 266 students during a module that has run predominately in virtual reality for the last three years, as well as the reflective experiences of the teachers.

KEYWORDS

Virtual Reality, The Metaverse, Higher Education, Student Experience

What is 'The Metaverse'?

The Metaverse is a term that originates in the 1992 book Snow Crash by Neil Stephenson referring to a future computer-generated universe *beyond* the physical world. Whilst precise definitions are currently lacking, there is an emerging consensus that several components are required for a true Metaverse (Dionisio et al., 2022), including: *Immersive realism* – users are psychologically and emotional engaged; *Ubiquity* – virtual worlds are accessible to all; *Interoperability* –users can move seamlessly between 'locations'; *Scalability* – capable of being accessed by whole populations.

When one considers the rapid rise during the Covid-19 pandemic in the maturing of immersive technologies, such as Virtual and Augmented Reality (VR/AR), together with more widespread access to social VR platforms – it is apparent that aspects of the Metaverse are already in place. Moreover, arguments have been made that a manifestation of the Metaverse is inevitable given the fact that humans evolved to comprehend the world through first-person experiences in spatial environments, rather than the flat 2D environments currently offered by computing and communication devices/systems (Dionisio et al., 2022).

Use of Virtual Worlds in Higher Education

A fundamental element to the Metaverse will be virtual worlds, defined as "*Shared, simulated spaces which are inhabited and shaped by their inhabitants who are represented as avatars*" (Girvan, 2018, p. 1099). In this respect, the education sector is widely seen as a potentially huge consumer of virtual worlds – enabling improved access of students to learning (from wherever in the globe), as well as enhanced flexibility in how/when learning is undertaken – Lee et al. (2021). Moreover, from a learning theory perspective, the widespread use of virtual worlds in education potentially provides numerous benefits, related to experientialism (learning while doing), socialisation (learning with others, including empathic understanding) and contextualization (learning in authentic/relevant environments) – Radianti et al. (2020).

In the Higher Education sector, there is emerging recognition that students need to be prepared better for their future digital working/leisure lives – by gaining a deep understanding of what it means to design, build and then be immersed and interact in virtual worlds (Lee et al., 2021). Nevertheless, previous Human Factors work related to virtual worlds in education has largely focused on one-off/solo experiences with immersive technology, usually from the perspective of training (discussed in Lee et al., 2021) - neglecting the complex social interactions that will arise when students and teaching staff interact as avatars across a long-time period.

The Nottopia experiment

At the University of Nottingham, a persistent, fantastical virtual world (referred to as Nottopia) has been used as the primary approach to teaching for a specific module for the last three years, initially conceived in 2020 at the outset of the Covid-19 pandemic but continuing beyond. The module is taken predominately by final year undergraduate and postgraduate taught students in Mechanical Engineering, Product Design, Human Factors and Ergonomics and Human-Computer Interaction (49 in 2020/2021; 95 in 2021/2022; 122 in 2022/2023) and concerns the Human Factors design issues for immersive technologies. Consequently, it is a perfect test-bed module to understand the social behavioural issues regarding students and teachers interacting over an extended period (three months of a semester) in virtual worlds. Nottopia can be accessed on desktop and mobile devices, but increasingly is being accessed via VR headsets either loaned to students or already owned by them. Teaching activities have included lectures (delivered either live in VR or pre-recorded) complimented by highly interactive seminars utilising a range of approaches, such as treasure hunts, design workshops, show and tells, 'fireside' chats, virtual field trips, etc. Sessions have been recorded (60 hours of video data) and students have been surveyed (40% response rate).

Five key guidelines for teaching in the Metaverse

Results from this research have already been reported extensively in Burnett, Harvey and Kay (2022). Here, we will summarise some key learnings, based partly on the formal data collected, but predominately on our reflections as teachers within this highly novel context. Subsequently, the following five basic recommendations/ guidelines are proposed of relevance to any practitioner considering the use of virtual worlds in their pedagogy:

1. *Contextualise* your virtual teaching space - immersing students in environments that are consistent with the topic under consideration to encourage creativity of thought.
2. Maximise the *3D capabilities* of virtual worlds - allowing students to engage with objects in new and empowering ways, e.g. by viewing from unique perspectives, resizing, etc.
3. Empower your students with *appropriate levels of freedom*. Virtual worlds afford freedom in movement (e.g. flying) - but universal access to these superpowers can distract from learning outcomes and need to be granted where relevant, and managed when of benefit.
4. Exploit the fact that virtual worlds can be *persistent and editable*. Unlike most real-world classrooms a virtual world teaching space can be available 24/7 for students and also available long beyond the running of a course, e.g. to demonstrate a cohort's work. It can also be editable either by a teacher or potentially by students – to be more relevant to what is being covered that week and/or encourage students to learn throughout a week.
5. Don't just revert to everything you do when teaching in the real-world. It is natural/easier to resort to traditional approaches to teaching, even in a virtual world (e.g. 2D lecture slides on a screen). Nevertheless, it is important ultimately to be creative and consider how the technology affords *radically new activities* of benefit to student motivation and learning.

References

Burnett, G.E., Harvey, C., & Kay, R. (2022) Bringing the Metaverse to Higher Education: Engaging University Students in Virtual Worlds. In A. Correria, V. Viegos, (Eds), *Methodologies and Use Cases on Extended Reality for Training and Education,* IGI Global Girvan, C. (2018). What is a virtual world? Definition and classification, *Education Technology Research Development*, 6, 1087–1100.

Dwivedi, Y.K. et al. (2022). Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy, *Journal of Information Management*, 66.

Lee, M. J. W., Georgieva, M., Alexander, B., Craig, E., & Richter, J. (2021). *State of XR & Immersive Learning Outlook Report 2021*. Walnut, CA: Immersive Learning Research Network.

Radianti, R., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020) A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda, *Computers and Education*, 147.

Anthropometric and ergonomic assessments of braiding activity among female hairdressers in Lesotho using the ART method

Tebello Pusetso[1], Samuel Mekonnen[1] & Hui Lyu[1,2*]

[1] International School of Design, Zhejiang University, [2] Ningbo Research Institute, Zhejiang University

SUMMARY

Hairdressers in their careers are at a high risk of work-related musculoskeletal disorders (MSD). The braiding process requires hairdressers to endure prolonged repetitive hand movements and awkward body postures. This study conducted hand anthropometric measurements of female hairdressers in Lesotho and assessed their risk of muscle injury during hair-braiding activities using the Assessment of Repetitive Task (ART) method. Results found that braiding activity is at high risk of muscular injury.

KEYWORDS

Braiding, Assessment of Repetitive Task, Anthropometry, Hairdressers

Introduction

Hair-dressing involves prolonged standing, repetitive hand movements, awkward trunk and upper-limb posture, and long working hours, which may contribute to discomfort, pain, fatigue, and musculoskeletal disorders among hairdressers (Ercan et al, 2022). According to the Korean Group for Occupational Medicine, 61% of hairdressers complained about shoulder pain, 59.9% of neck pain, 53% of lower back pain, and 42% of hand and wrist pain (Sy et al, 2016). Braiding is the interlacing of two or more strands at an angle (Talha et al, 2021), which is especially popular among African women. However, various braiding styles take between 3.5-14 hours to complete depending on the client's head size, hair type, and volume. Moreover, a braider must execute more than 50 repetitive wrist and finger motions in 1 minute, increasing their discomfort rate in the fingers and wrist/hand. Therefore, this study aims to capture hairdressers' activities in Maseru, Lesotho; measure their hand anthropometric dimensions to provide a reference for automatic braider design; and assess the braiding process using the ART method to determine the risk of muscle injury.

Methodology

Forty-one female hair-dressers, aged from 21 to 46 years old, were recruited from 14 different hair salons in the Maseru, Lesotho. Their hand anthropometric dimensions were measured using a flexible measuring tape and a plastic ruler. Among them, nineteen subjects were further observed and interviewed about their hair salon activities, and one of them was randomly selected to be assessed throughout the braiding process using the Assessment of Repetitive Task (ART) tool (HSE 2010).

Results

Hand anthropometric dimensions were shown and compared with Pheasant (2003) and Ching-yi and Deng-chuan (2017) in Table 1.

Within a day of observation, we found that 17 hairdressers received 1-5 customers for hair braiding, and two hairdressers received 6-10. To complete a large braid, 17 hairdressers took 1-2 hours, and two hairdressers took 3-4 hours. To complete a medium braid, ten hairdressers took 1-2 hours, and nine hairdressers took 3-4 hours. To complete a small braid, all the hairdressers reported a 3–4-hour work.

For the ART results (see Table 2), we assessed the whole braiding process for 3 hours and 45 minutes, which includes 10 minutes of hair washing, 15 minutes of hair drying, and 3 hours and 20 minutes for braiding. Results indicated that hair washing is at a moderate-risk level, hair drying at a low-risk level, and braiding is the most hazardous task at a high-risk level for both sides.

Table 1. Hand anthropometric dimensions of female hairdressers in Lesotho

	This study				Pheasant (2003)	Ching-yi and Deng-chuan (2017)
Hand dimensions(mm)	**Min**	**Max**	**Mean**	**SD**	**Mean**	**Mean**
Hand length	158	188	173	8.5	174	167.9
Palm length	95	117	105	5.6	97	-
Thumb length	51	67	56	4.6	47	
Index finger length	59	72	66	3.9	67	-
Middle finger length	60	82	76	5.8	77	-
Ring finger length	60	78	68	3.9	66	-
Little finger length	48	62	54	3.7	50	-
Thumb breadth	17	25	21	2	19	-
Index finger breadth	12	18	15	1.9	18	-
Hand bredth (metacarpal)	72	82	77	2.7	76	75.2
Hand breadth (across thumb)	87	96	91	2.5	92	-
Maximum spread	162	189	176	6.5	190	-
Grip breadth inside width diameter	36	47	41	2.2	48	38.3
Grip breadth inside length diameter	27	31	31	1.9	-	30.7

Table 2. ART results of hairdressing work activities

	Hair washing		Hair drying		Braiding	
	Left	Right	Left	Right	Left	Right
Task score	24	24	22	23	33	33
Duration multiplier	0.5	0.5	0.5	0.5	0.75	0.75
Exposure value	12	12	11	11.5	24.75	24.75
Risk	Moderate	Moderate	Low	Low	High	High

Conclusion

The ART results confirm that hairdressers are at a high injury risk during braiding due to awkward postures and repetitive upper limb movements. Therefore, a new design of an automatic braiding machine with an ergonomically-design handle will be suggested to create a more comfortable braiding process with faster speed and fewer injuries for hairdressers.

References

Ching-yi, W. and Deng-chuan, C. (2017). Hand tool handle design based on hand measurements. EDP Sciences. DOI: 10.1051/matecconf/2017119010.

Ercan, S., Parpucu, T., Başkurt, Z. and Ferdi Başkurt. (2022). Gender differences, ergonomics risks and upper quadrant musculoskeletal pain in hairdressers. Internation Journal Of Occupational Safety And Ergonomics (JOSE). Doi:10.1080/10803548.2022.2066315.

HSE. (2010). Assessment Of Repetitive Tasks Of The Upper Limbs (The ART Tool) Guidance for Employers, England: Health and Safety Execution (HSE).

Sy, O., & Phillips, M. (2016). Musculoskeletal Symptoms and Associated Risk Factors Among African Hair Braiders. Journal of occupational and environmental hygiene, 13(6), 434-441.

Talha I., Syed Zoraiz H., Syed Murtaza A. and Usama Bin A. (2021). Automatic Braiding Machine-Its Design and Analysis. DOI: 10.13140/RG.2.2.22543.30881.

Can Intersectionality Increase Active Travel in Marginalised Groups? A Literature Review

Joy McKay[1], Katie Parnell[1] & Katie Plant[1]

[1]University of Southampton, UK

SUMMARY

Active travel, such as walking, running and cycling, are cheap, sustainable, and healthy ways to transit within in the urban environment. Many marginalised groups are either underrepresented in active travel modes or find they are limited using them in certain neighbourhoods or at certain times. These limitations lessen the accessibility of a range opportunities including those of employment, social activities and cultural experiences to a wide range of citizens. This review endeavours to recognise those barriers to active travel which affect a diverse selection of society and understand affordances which encourage use of these modes. It aims to identify solutions which may encourage active travel across a diverse community leading to an urban environment which is more equitably accessible for all.

KEYWORDS

Gender, Active Travel, Running, Cycling, Walking, Intersectionality

Introduction

Intersectionality is a term first used to describe identity by feminist Kimberle Crenshaw. She proposed that identities are constructed through the crossroads where multiple dimensions of the individual meet. Initially this was in relation to the increased levels of oppression experienced by black women due to the intersection of their race and gender but she recognised this analytical approach could be extended to other “marginalisations” (Crenshaw 1991). Following this an intersectional lens has been applied to research across a diverse range of domains including transportation research (O’Brien 2020). These intersectional groups are now often extended to include any combination of gender, race, sexuality, transness, culture, religion, age, class, education, body shape, nationality, language, immigration status, occupation, and disability. (Lim et al 2021, Stanley 2020).

In the UK 42% of women and 34% of men do not meet activity targets (WHO 2020) resulting in an estimated annual cost of £7.4 billion and is thought to be the leading cause of one in six deaths. (Office for Health Improvement and Disparities 2021). By including active travel in the commuting routine improvements can be seen to health, air quality, congestion and carbon footprints (Nieuwenhuijsen 2021).

Aims and Approach

According to intersectional theory whilst gender clearly plays a major role in the accessibility of active travel this single aspect of the individual should not be taken in isolation (Francis and Pearce, 2020). This review endeavours to identity, from the current literature, those barriers to active travel which affect a diverse range of members of society.

Method

A Grounded Theory was taken when conducting this literature search. As such no pre-determined categories of classification were derived before reading the literature, instead the themes emerged from the literature, evolving throughout the reviewing and re-reviewing process. This approach has the benefit of removing bias based on previous topic knowledge, or reflexivity, allowing the researcher to approach the literature without an existing hypothesis in order that ideas arise inductively (McGhee et al 2007).

The key term 'intersectionality' was combined with either 'active travel', 'walking', 'cycling' or 'running'. Web of Science and Scopus were selected for identifying published academic literature in addition Google Scholar was used to ensure grey literature and policy documentation, this was due to the still emerging use of intersectionality within the transport domain and the importance of its inclusion in not just academia but also in transport policy and decision making. Active travel modes, marginalised groups, and barriers or incentives relating to engaging in active travel were thematically categorised.

Results

A final selection of 41 papers were identified. These come from a wide range of domains including Human Factors, Health Promotion, Sociology of Sport, Urban Design, Climate Change, Transport Planning and Geography. 28 papers mentioning cycling 21 mentioned walking (several focusing on both) but running was under represented being mentioned in just two papers.

The marginalized groups identified, with their frequency were Gender (39), Race (23), Age (starting at 55+) (19) and Socio-Economic-Status (SES) (19). LGBTQ+ (19). As gender was the most commonly occurring group, it therefore intersected with the other groups most frequently. Intersecting with race 22 times, with age 18 times, SES 19 times and LGBTQ+ 19 times. Most common barriers identified were Traffic Safety (18), Personal Safety including experiencing microaggressions, bullying and fear of sexual assault (14). Each scoring four are travelling with children, the limited range of the mode versus required distance to travel, racism and exposure to pollution. The literature focusses more on barriers however the most common incentives were Social benefits of being with others or seeing friends/ family (7) Good Infrastructure (4) and improvements to Weight or Fitness (4).

Discussion

Safety was by far the most common barrier both from traffic, and personal safety, fear of gender specific abuse or attack. Whilst improved cycle infrastructure was a solution to the fears surrounding traffic safety (Russell et al 2021, Graystone et al 2022) those surrounding personal safety are a more complex societal issue. Related to this concerns about how women engaged in active travel are perceived as appearing unfeminine, putting themselves before familial commitments (Curry 2016) or simply a feeling of not belonging in the domain of the straight, middle-class, white Cis man (Adjepong, 2022, Stanley 2020). Some of these issues can be resolved through belonging to targeted groups (Wegner et al, 2019, Ravensbergen et al 2020) and being able to identify with those who are depicted as taking part in active travel (Fogg-Rogers et al 2021).

Conclusion

A literature review was conducted to determine which marginalised identities intersected with gender when creating barriers to active travel update and whether there had been incentives previously identified to counteract this. Safety was identified as the biggest barrier across all groups whereas social benefits were seen as the biggest incentive. Future work will look further into the reasons for the barriers to engaging in active travel in order to identify ways in which they can been removed.

References

Adjepong, A. (2022) 'We're, like, a cute rugby team': How whiteness and heterosexuality shape women's sense of belonging in rugby. International Review For The Sociology Of Sport

Crenshaw, K. (1991) Mapping the Margins: 'Intersectionality, Identity Politics, and Violence against Women of Color' Stanford Law Review

Curry, W. B. (2016) Health and physical activity messages among ethnic minority groups in Dagkas, S & Burrows, L (Eds) Families, Young People, Physical Activity and Health, Routledge

Fogg-Rogers, L., Hayes, E., Vanherle, K., Pápics, P. I., Chatterton, T., Barnes, J., Slingerland, S., Boushel, C., Laggan, S., & Longhurst, J., (2021) Applying Social Learning to Climate Communications—Visualising 'People Like Me'in Air Pollution and Climate Change Data. Sustainability

Francis, S. and Pearce, K. (2020) Reimagining movement and the transport appraisal process through a gender lens: a case study in the UK utilising a lifecycle approach. Jacobs Consultancy UK

Graystone, M., Mitra, R. & Hess, P. M. (2022) Gendered perceptions of cycling safety and on-street bicycle infrastructure: bridging the gap. Transportation research part D: transport and environment

Lim, H., Jung, E., Jodoin, K., Du, X., Airton, L. & Lee, E. (2021) Operationalization of Intersectionality in Physical Activity and Sport Research: A Systematic Scoping Review. SSM Population Health

McGhee, G., Marland, G.R. & Atkinson, J.M. (2007) Grounded theory research: literature reviewing and reflexivity. Journal of Advanced Nursing 60(3):pp. 334-342. Stanley 2020

Nieuwenhuijsen MJ. (2021) New urban models for more sustainable, liveable and healthier cities post covid19; reducing air pollution, noise and heat island effects and increasing green space and physical activity. Environ Int. 2021

O'Brien, K. (2019) Planning Sustainable Public Transport Using an Intersectional Lens http://meetingoftheminds.org/planning-sustainable-public-transport-using-an-untersectional-lens-30926 (accessed 2/3/20)

Office for Health Improvement and Disparities (2021) Physical activity: applying All Our Health https://www.gov.uk/government/publications/physical-activity-applying-all-our-health/physical-activity-applying-all-our-health (accessed 20/12/22)

Ravensbergen L (2022) 'I wouldn't take the risk of the attention, you know? Just a lone girl biking': examining the gendered and classed embodied experiences of cycling. Social & cultural geography

Russell, M., Davies, C., Wild, K. & Shaw, C (2021) Pedalling towards equity: Exploring women's cycling in a New Zealand city. Journal of transport geography

Stanley, P. (2020) Unlikely hikers? Activism, Instagram, and the queer mobilities of fat hikers, women hiking alone, and hikers of colour. Mobilities

Wegner, C. E., King, C & Jordan, J. S. (2019) The role of organizational membership in overcoming dissonant sport activity identities. Sport Management Review

WHO (2020) Who Guidelines on Physical Activity and Sedentary Behaviour. Geneva: World Health Organization

Assessing the Benefits of Virtual Reality Training for Manual Assembly

Charlotte Temmink[1] & Yee Mey Goh[1]

[1]Loughborough University

SUMMARY

The aim of this study is to assess, through experimentation, the effectiveness of Virtual Reality training when compared to a traditional written instructions in a manufacturing assembly process. The effectiveness was evaluated in terms of perceived workload and task performance, with participants asked to self-report on the experience during both the written and VR training. Findings indicate that VR training reduced the user's workload perception and errors made during the task, however there was no significant impact on time taken to complete the task.

KEYWORDS

Virtual Reality, Assembly, Workload

Introduction

Virtual Reality (VR) technology has seen rapid advancement in recent years, where applications for training can be found across many industries. However, literature in this area gives conflicting views on whether VR training promotes greater knowledge uptake and retention (Makransky, 2021), with effects of VR on physical and mental workload remaining unevaluated (Werk, 2021). This study (part of an undergraduate student project) investigates the effectiveness of VR training when compared to a traditional written method. Assessment was conducted through experimentation, using the example of the assembly process of a Wankel engine.

Method

The VR program involves an interactive virtual environment that allows the user to manipulate the assembly and parts through 360 degrees of rotation, developed through Unity and Steam VR giving instructions and tips to the user at every stage. The number of errors and time taken to assemble the engine were collected and participants in both groups self-reported their perceived workload after performing the tasks using the NASA Task Load indeX (TLX) (NASA, 2020). The participants were recruited to represent equal sample sizes between gender and those with and without an engineering background to reflect different experience level with assembly tasks. Following ethics and risk assessment, participants with motion sickness or pregnancy were excluded.

Results

The data was evaluated and analysed to establish whether there is statistical significance as to whether Virtual Reality Training improved performance (decrease in assembly time and errors) and reduced the perceived workload of the participants during the task. 22 samples were collected as part of the experiment and fixed ANOVA t-test were performed. Group 1 (control group) performed the assembly with written instructions first, then the VR training. Group 2 (experiment group) performed the VR training first then the assembly with written instructions. Participants completed the NASA-TLX questionnaire after completing each task as shown in Figure 1.

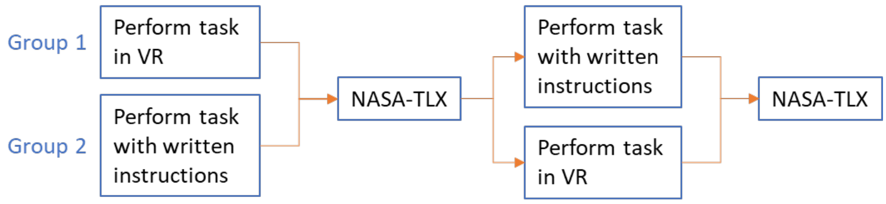


Figure 1: Experimental design (Group 1, n=11; Group 2, n=11)

a) Performance

There was no decline in assembly time between the groups (i.e. learning effects). The mean completion time did not significantly improve with VR training, however the variance is reduced, p=0.019 (Figure 2a). The number of errors was reduced with VR training, p=0.137 (Figure 2b), with participants making fewer errors locating parts.

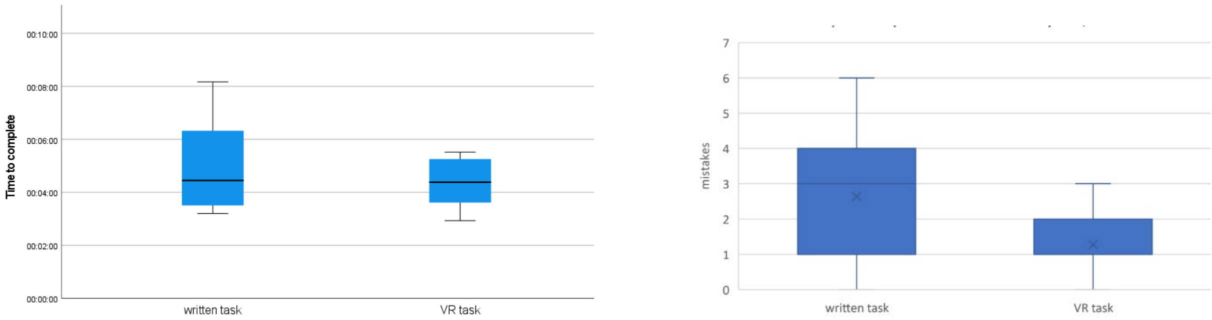


Figure 2: Boxplot of a) time to complete assembly; b) mistakes made during assembly.

b) Workload perception

The average workload for both groups was lower when assessing the VR task compared to the written task (Figure 3). For Group 1, the biggest change in workload was seen in the *temporal demand*, and the smallest change in workload was observed in *frustration*. For Group 2, the biggest change in workload was seen in *time demand*, but the smallest was seen in *performance*.

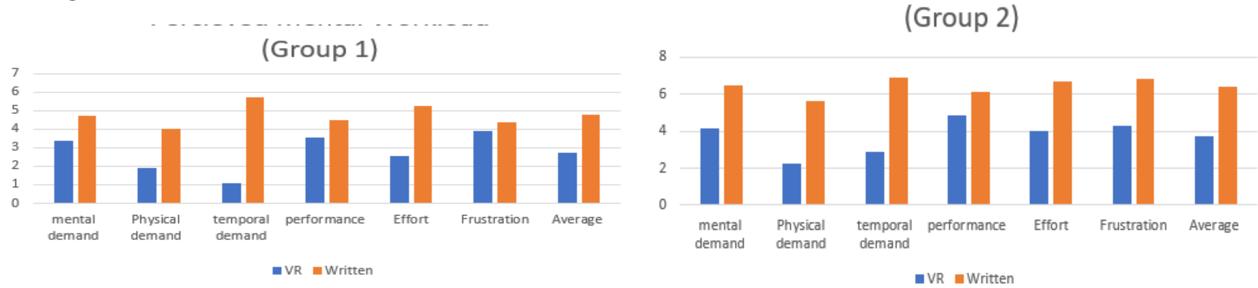


Figure 3: Perceived workload a) Group 1; b) Group 2.

Conclusions

The results of the experiment indicate that VR training did improve the user's perception of workload and reduce the number of errors made in the assembly task. However, there was no significant impact on time taken to complete the assembly task. Perceived workload was reduced in almost all NASA-TLX categories and weighted average workload was significantly lower in VR compared to the traditional written method. Whilst there was an overall improvement in performance observed in the participants who were given the VR training prior to the assembly task, there were some areas where there was little improvement over the control group. These errors include the incorrect fit of the drive shaft or rotor and could be due to the lack of haptic feedback.

References

Makransky, G., Andreasen, N. K., Baceviciute, S. & Mayer, R. E. (2021) Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. Journal of Educational Psychology, 113(4), 719–735.

NASA. (2020) NASA TLX: Task Load Index. https://humansystems.arc.nasa.gov/groups/tlx/. Accessed: 12/01/2022.

Werk, L. P. & Muschalla, B. (2021) Workplace mental health promotion in a large state organization: Perceived needs, expected effects, neglected side effects. Open Res Europe 2021, 1:17

Voluntariness and extent of telework – association with heart rate variability

Leticia Bergamin Januario, Marina Heiden, Svend Erik Mathiassen, Tea Korkeakunnas, Gunnar Bergström & David M Hallman

University of Gävle, Sweden

SUMMARY

We evaluated the impact that (1) voluntariness and (2) extent of telework had on 24h heart rate variability (HRV) measured objectively for three days in white-collar workers during the COVID-19 pandemic. Workers with high voluntariness had higher root mean square of successive differences between R-R intervals (RMSSD), however high extent of telework did not affect HRV metrics, after adjusting for suitable covariates. These results may indicate higher parasympathetic activity, an indicator of good health.

KEYWORDS

24h time-use, remote work, telecommuting, working from home.

Introduction

Telework increased significantly during the COVID-19 pandemic and telework became normative in many white-collar occupations. However, little is known about the effects of telework voluntariness on psychophysiological responses, and whether those effects depend on telework extent (Bouziri et al. 2020). The aim of this study was to investigate associations of voluntariness and extent of telework with 24-h measures of HRV, during the COVID-19 pandemic.

Methods

We used data from the *Flexible Work: Opportunity and Challenge* cohort (Svensson et al. 2022), including 294 white-collar workers from 8 companies, working ≥50% of full time, who had answered a web-survey and accepted to participate in technical measurements of heart rate monitoring and physical activity (PA).

Data collections and analysis

Telework was measured using two questions, one about voluntariness to telework ("Do you currently have the freedom to choose whether you want to do your work remotely or not?") and another about the extent of telework ("How much do you use the opportunity to work remotely during the ongoing pandemic?"). The possible answers were categorized as either high ("To a very high degree / quite a lot") or low ("To some extent / not at all'"). Each worker participated in technical measurements initiated by the research team at the workplaces or during on-line meetings, where body height and weight were also assessed to calculate body mass index (BMI). A heart rate monitor was used for three days (Bodyguard2, Firstbeat Technologies Ltd., Jyväskylä, Finland) and data were processed in the Acti4 software (Skotte and Kristiansen 2014). Heart rate (HR), RMSSD and standard deviation of R-R intervals (SDNN) were derived during work, leisure, and sleep, as identified through a diary. A thigh-worn accelerometer (Axivity AX3, Axivity Ltd, Newcastle, UK) was used to assess PA for 7 days in terms of moderate to vigorous PA and sedentary/ light PA behaviors, using validated algorithms (Skotte et al. 2014) according to compositional data analysis (Hallman et al. 2021). Each company provided information about the workers' age and gender. We ran unadjusted and adjusted multilevel linear mixed models to estimate the effects of voluntariness and extent of telework on HRV, considering the period of the day (work, leisure, and sleep). Adjusted models included age, gender, BMI, and PA level. We performed all tests in SPSS (v. 27, IBM Corp, Armonk, NY, USA) with significance level at 0.05.

Results

Telework voluntariness was significantly associated with RMSSD in both the unadjusted (not shown) and adjusted models (table 1). RMSSD reflects the variance in the beat-to-beat HR and is used to estimate changes in HRV that are mediated through the vagal system, which can indicate that the high voluntariness of telework is associated with higher HRV, even though effect sizes were small. Contrary to expected, the relationship between telework voluntariness and HRV was not affected by telework extent nor period of the day, since we found no interaction between voluntariness and extent of telework nor period of the day. A low extent of telework per se did not affect HRV metrics after adjustment for covariates. Work and leisure were both associated with increased HR and reduced RMSSD and SDNN compared with sleep. Higher values of RMSSD and SDNN can be translated to efficient autonomic mechanisms (particularly parasympathetic cardiac modulations), which indicates good general health.

Table 1: Estimated effects (p-values) on heart rate (beats-per-minute - bpm), RMSSD (milliseconds - ms) and SDNN (ms) for high and low degrees of telework voluntariness and extent, as well as for different periods of the day.

		Heart rate (bpm)	**RMSSD (ms)**	**SDNN (ms)**
Voluntariness of telework	Low	0.55 (0.40)	**-3.32 (0.01)**	-1.74 (0.17)
	High	0.00[a]	0.00[a]	0.00[a]
Extent of telework	Low	-0.08 (0.90)	1.35 (0.33)	1.81 (0.19)
	High	0.00[a]	0.00[a]	0.00[a]
Period of the day	Leisure	**16.40 (<0.01)**	**-15.19 (<0.01)**	**-6.93 (<0.01)**
	Work	**14.11 (<0.01)**	**-13.68 (<0.01)**	**-4.05 (0.01)**
	Sleep	0.00[a]	0.00[a]	0.00[a]
Models adjusted for age, gender, physical activity (accelerometry ILRs) and BMI. a. reference category. Bold values represent statistical significance				

Conclusions

This study showed that during the COVID-19 pandemic, workers with high voluntariness of telework had higher parasympathetic indicators of HRV than those with low voluntariness. Thus, it seems that the voluntariness to choose where to work (at the office or at home) during the pandemic had small, but beneficial effects on parasympathetic activity, which may be relevant to worker health. Telework extent did not affect the HRV metrics evaluated, however, it is important to acknowledge that in this study we measured telework extent as in the opportunity to telework "to a very high degree" or "not at all", and not as the frequency of telework in hours or days.

References

Bouziri, Smith, Descatha, et. al., 2020. "Working from Home in the Time of COVID-19: How to Best Preserve Occupational Health?" OEM 77(7):509–10.
Hallman, Januario, Mathiassen, et.al., 2021. "Working from Home during the COVID-19 Outbreak

in Sweden: Effects on 24-h Time-Use in Office Workers." BMC Public Health 21(1):1–10.
Skotte and Kristiansen. 2014. "Heart Rate Variability Analysis Using Robust Period Detection." BioMed Eng OnLine 13:138.
Skotte, Korshøj, Kristiansen, et al., 2014. "Detection of Physical Activity Types Using Triaxial Accelerometers." J Phys Act Health 11:76–84.
Svensson, Hallman, Mathiassen, et. al., 2022. "Flexible Work: Opportunity and Challenge (FLOC) for Individual, Social and Economic Sustainability…" BMJ Open 12(7):e057409

Assessing pilots' mental workload using touchscreen inceptor for future flight deck design

Joao Paulo Macedo, Kyle Hu, Rani Quiram & Samarth Vilas Burande

Safety and Human Factors in Aviation MSc, SATM, Cranfield University, United Kingdom

SUMMARY

Touchscreen displays are one of the pillars of future flight deck design and it is foreseen that at some point traditional flight control inceptors will be modified to a touchscreen version. However, this transition can only be safe and successful with due regard for human performance implications. This study addresses it by comparing pilots' mental workload for a traditional sidestick and an innovative touchscreen control inceptor. The results indicate that the new technology increases pilot workload, suggesting that further development is required to use it in future flight decks.

KEYWORDS

Cockpit design, Human-machine interface, Mental workload

Introduction

Touchscreen displays are already featured in the cockpit of commercial aircraft and the use of this technology is proposed for the next generation of flight decks (Li et al., 2022). Since touchscreens have been proved beneficial for certain applications (Stanton et al., 2013), futuristic cockpit designs explore the possibility of replacing traditional flight controls with touchscreen inceptors. While such a replacement could reduce pilot training time and costs, a new control inceptor must address all requirements for manual flight control and be compliant with human-centered design principles. To understand the characteristics of an innovative flight control inceptor from a human factors perspective, this paper investigates differences in pilots' mental workload when using a traditional sidestick and a touchscreen inceptor in a realistic operational environment.

Methods

The experiments involved 26 participants aging from 21 to 46 years old (M = 28, SD = 7.1) with varying levels of flight experience (M = 257.7, SD = 709.3) and were conducted in the Future Systems Simulator (FSS). The FSS is a reconfigurable simulator, allowing the quick switch between traditional flight controls to an innovative touchscreen inceptor consisting of a touch sensitive region in the Primary Flight Display (PFD) in which a sign can be moved to control the attitude of the aircraft. Each participant executed a landing scenario using both inceptors. Perceived workload was measured using NASA-TLX, which consists of six load dimensions to be rated from 0 to 100 (Hart & Staveland, 1988).

Results and Discussion

Figure 1 shows the mean NASA-TLX scores for both inceptors. Two-tailed paired t-tests for sample means showed a significant effect of the inceptor in perceived mental workload (NASA-TLX total score) – $t(25) = 2.0$, $p < 0.05$, $d = 0.40$ –, more mental demand – $t(25) = 2.6$, $p < 0.005$, $d = 0.50$ – and frustration – $t(25) = 3.31$, $p < 0.05$, $d = 0.65$. The comparisons for all other partial scores were not statistically significant. The results indicate that the innovative technology increases perceived workload, requiring more mental resources and resulting in higher frustration compared to a traditional sidestick. It is hypothesised that these differences are a consequence of the design and location of the new inceptor. Using the touchscreen control makes it more difficult for the controller to capture relevant information to perform the task (e.g., attitude, glideslope indicator), since the PFD is partially blocked by the controller's own hand, creating the need to look at more places and integrate information from different sources (e.g., outside of the window). Moreover, the lack of a neutral position for the control, associated with the absence of control force feedback and physical barriers to indicate control stops, requires the controller to look at the position of her/his own finger to assimilate command inputs. In addition, the pitch axis in the touchscreen case is reversed in relation to traditional controls, generating an extra mental demand. This combination of factors makes the task more complex and stressful to be executed using the touchscreen inceptor, resulting in an overall higher mental workload.

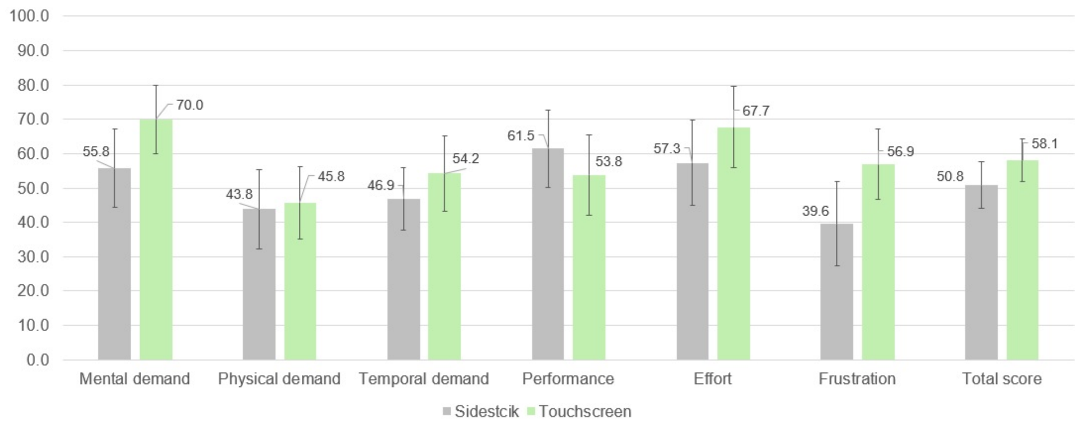


Figure 1: NASA-TLX scores for touchscreen and sidestick

Conclusion

This study suggests that replacing a traditional flight control inceptor with a touch screen increases pilot mental workload, which is linked to the pilots' lack of familiarity with the new design and information seeking. Although touchscreen technology has been proved as beneficial to other applications in the flight deck, it should still be improved as a flight control inceptor to comply with human-centred principles and meet pilot control needs in consideration of the future cockpit design.

References

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (Vol. 52, pp. 139–183). North-Holland.

Li, W. C., Liang, Y. H., Korek, W. T., & Lin, J. J. H. (2022). Assessments on Human-Computer Interaction Using Touchscreen as Control Inputs in Flight Operations. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *13307 LNAI*, 326–338.

Stanton, N. A., Harvey, C., Plant, K. L., & Bolton, L. (2013). To twist, roll, stroke or poke? A study of input devices for menu navigation in the cockpit. *Ergonomics*, *56*(4), 590–611.

Human factors in emergency management

Weixuan Li, Glyn Lawson & Gary Burnett

Human Factors Research Group, Faculty of Engineering, University of Nottingham, UK

SUMMARY

Good emergency management is essential when human beings face natural disasters. However, there are many human factors challenges in this area, in particular teamwork amongst those involved with the response efforts, but also communication, navigation, and workload. This article presents the findings from nine interviews with emergency management practitioners based in China, with an emphasis on the potential use of Virtual- and Augmented-Reality to address the aforementioned human factors issues.

KEYWORDS

Emergency management, Virtual Reality, Augmented Reality.

Introduction

Natural and artificial disasters such as earthquakes, floods, and workplace incidents are all crises that humankind must deal with. Take, for example, flood: according to UK government figures, almost 5 million people in England and Wales live in flood-prone locations. However, 25% of flooding occurs outside areas formally designated as being flood-prone. Annual flood damage costs are £1.1 billion across England (EFRA, 2021; GOV.UK, 2022). Emergency management plays a critical role in reducing the impact and loss of life and economy in these disasters (Wilson & Oyola-Yemaiel, 2001). Emergency management can include disaster prevention, emergency preparation (training and safety planning), emergency response (evacuation and rescue), and disaster recovery (recovery of basic services such as hospitals and other life rescue services) (Murphy, 2007). Researchers and professionals will get a better understanding of the most appropriate response strategies necessary for diverse catastrophes via emergency management research, therefore enhancing emergency management measures and creating a safer environment for the public.

Emergency management teams face numerous human factors challenges, for example, emergency management is highly dependent on teamwork (Hayes, 2017).. When responding to an emergency, emergency management teams are often faced with issues such as task allocation, team decision-making, team communication, team leadership and trust between team members (Frye & Wearing, 2016; Johnson, 2017; McLennan et al., 2017; Owen, 2017). In natural disasters, emergency management teams are often confronted with complex environments that are unpredictable and accompanied by a variety of unexpected factors, such as sudden secondary disasters or communication instability (Frye & Wearing, 2016). This usually has the effect of overloading the cognitive capacity of emergency responders. The cooperation of team members will determine whether the team is able to respond to the complex disaster environment successfully (Hayes, 2017), which is why good teamwork is so essential in emergency management.

Communication is an issue that most teams face, but this can be even more significant for emergency management teams (Hayes, 2017). Emergency management teams often work in a variety of disaster scenarios and are exposed to complex disaster environments that greatly affect team communication. Especially in the rescue of large natural disasters such as earthquakes and floods, the interruption of power and the damage of communication facilities have greatly affected the communication ability of rescue teams. (Fischer III, 1996). Good team communication and management will greatly enhance the effectiveness of emergency management teams. Natural disasters can present a particularly large challenge to communication, as they often require a coordinated response from several teams. Cooperation between different teams relies on communication ability of the team leader. (Kapucu & Hu, 2016). Furthermore, in an emergency management network, communication between government agencies and public and private organisations can also have a significant impact on the response (Kapucu & Garayev, 2011). The emergency rescue team should also have efficient methods of communication within team members. (Owen, 2017). How to improve the effectiveness of communication in emergency management networks is, therefore, an issue that requires more research and input.

In emergency situations, stress is an important factor in emergency decision-making. A study by McLennan et al. (2017) shows that in wildfire rescue, firefighter stress often leads to difficulties in decision-making and rescue operations. Stress can have a large impact on memory, decision making, attention, and perceptual-motor skills. This is why training before an emergency and reviewing after an emergency are so important. These actions will help emergency teams reduce the impact of stress during emergencies, especially in training for worst case scenarios (Johnson, 2017). Considering emergency training, worst-case scenario simulations are the most frequently used training method (Johnson, 2017). By simulating the worst-case scenarios, emergency response teams are able to improve their effectiveness and teamwork in real-life disaster situations. In future research, how to conduct simulation training through VR and AR devices will be an important research direction. These studies will help the emergency rescue team to alleviate stress and other human factors problems they face in the rescue, to help them better complete the decision-making in the rescue work.

A severe natural disaster is a very complex environment that involves many people and circumstances, such as the emergency management team, the residents of the affected area, the businesses in the affected area, etc. Frye (2017)'s research on forest fires shows that fire commanders often use metacognitive skills (cognitive models based on previous disaster response experience) to help them think under complex cognitive loads. Metacognitive helps commanders to make better decisions and thus reduce the damage caused by fires. Thus improving the cognitive abilities of emergency response teams can help them to better cope with emergencies during disasters.

In summary, in addition to teamwork challenges, emergency team members are often faced with a number of human factors challenges which may be experienced individually, even if working in a team, such as stress, workload, cognitive, decision making, and excessive demands on working memory (Owen, 2014) These human factors problems also have a significant impact on the cooperation and efficiency of emergency management teams (McLennan et al., 2017). How to address these issues is, therefore, a key consideration for researchers.

Methodology

I designed a semi-structured interview to confirm that the human factors issues reported above in the academic literature are those encountered by front-line emergency rescue workers in real-world rescue scenarios, and to explore in greater depth the nature of these issues from the perspective of these rescue workers. Ten emergency management practitioners from Sichuan, China, were interviewed, including firefighters, emergency management personnel, members of civilian rescue teams, and others. The sample size was restricted to ten participants as this is a specialist population, who need government approval to participate, so the overall population is not large nor easily accessible. One participant withdrew from the interview thus the analysis is based on the interview data of nine participants. The interview was approved by University of Nottingham's Faculty of Engineering Ethics Review Committee.

The first set of questions in the interview focused on learning about the participants' work backgrounds in order to obtain an in-depth understanding of their experience. We discussed their work over the last few years to decades, with a focus on emergency rescue work. The second set of questions in the interview focused on the human factors areas identified in the previous literature (Table 1). The goal was to determine whether the problems identified in these documents exist in first-line rescue work, as well as to establish the perspectives of front-line rescue teams on these issues. These issues are primarily associated with teamwork, communication, workload, navigation, and work stress.

Table 1: Interview questions about human factors challenges identified from previous literature.

HF issues	Questions
Teamwork	What is your current teamwork model? What do you think is the biggest problem in the existing team cooperation? What works well?
Communication	What is your communication mode in the current team cooperation? How do you complete the communication between the rear support team and front-line staff at the emergency front? Where are you during this process? Are there any problems/issues with the current communication mode? What works well?
Workload	Do you experience a large amount of complex information? Does this impact your workload in anyway? Please tell me about this.
Navigation	Have you ever experienced any problems with navigation during disaster areas? How do you navigate? What tools do you use to help you navigate? Do they pose any issues? What works well?
Stress	Do you experience any stress at work? If so, please explain how this manifests. How do you cope with any stress?

During my doctoral research, I intend to use Virtual Reality (VR) and Augmented Reality (AR) technology to assist rescue teams in resolving their human factor issues. As a result, I included the third set of questions in this interview (Table 2) to explore the perspectives and recommendations of front-line emergency staff regarding the use of VR and AR in emergency rescue. I showed participants current applications of VR and AR technologies in the field of emergency management prior to this group of questions. These applications included, for example, VR technology being used for fire safety training. I also showed Head Up Display (HUD) technology, which similar to AR technology, has been widely used in special forces. This technology can improve communication efficiency among special forces members and can help members obtain mission information faster (Goldiez et al., 2007). Furthermore, if AR technology is used in emergency rescue teams, such as post-disaster search, rescue and survey, it can theoretically improve the efficiency of communication and task execution (Park et al., 2018). The interviewees were asked about their attitudes towards applying AR and VR technologies in emergency management.

Table 2: Questions to understand interviewees' views on the application of new technologies.

Serial	Questions
1	What is your attitude towards using VR and AR technology in emergency management? Could you explain why you have such an attitude?
2	Through our introduction and your understanding of VR and AR. Do you think this technology could help you in your work? Please explain your answer.
3	Do you have any suggestions for using similar technologies in emergency management?

After the interview, I transcribed the interview recording, transcribed it into text and translated it into English. Then I used NVivo to code the topic of the interview and conduct qualitative analysis.

Results

Experience

All nine participants have rich experience in emergency rescue. Among the nine participants, one from the civilian rescue team had the least amount of emergency rescue work. His main line of work is as a lawyer, but in his spare time he participates in emergency rescue training and follows the rescue team to assist in disaster relief. One of the participants who worked the longest was a former fire commander with over 30 years of experience in the fire department.. He also assisted in the rescue of many large-scale natural disasters during his career.

Teamwork

Participants reported that in previous emergency situations, teamwork was the most important factor influencing the success of the rescue. Participant 2 is a firefighter and has led teams in earthquake rescue many times, most notably during the Wenchuan earthquake in 2008. Through close collaboration, he and his team completed many rescue tasks. The Chinese fire brigade was a paramilitary organisation at the time. Therefore, they placed a high value on teamwork training during their daily training process. As a result, Participant 2 reported having a strong team understanding of each task. However, Participant 2 also discussed the challenges of teamwork in large-scale disasters. For example, due to the smoke and fire, it can be difficult for them to confirm the position of their teammates during the search and rescue operations of a large fire incident (overlapping with *Communication* and *Navigation* issues). Simultaneously, due to noise interference on the scene, it is sometimes impossible to receive assistance requests from teammates from their walkie-talkies in a timely manner (also *Communication*). Participant 4 works for a public welfare rescue organisation. He has also assisted in flood rescue numerous times. In terms of teamwork, he claims that because they are from non-governmental public welfare rescue organisations, their training is often not as extensive as that of professional rescue teams. Participant 4 reported that many rescue operations have poor teamwork due to team members' lack of awareness of disaster relief operations or the urgency of the task. During my interviews, all nine participants mentioned similar issues to those raised by Participant 4. The degree of difficulty encountered in Teams differs due to the distinction between professional and ordinary training.

Communication

All nine participants reported that communication is a major challenge for emergency response teams. Despite the fact that relevant technologies and equipment are constantly being updated and iterated, communication difficulties were reported as frequently encountered in real disaster environments. Participant 1 has worked in safety supervision and has assisted in the rescue of many safety production accidents. He stated that communication is often the most difficult issue for rescue teams in coal mine accidents, as due to the obstruction of underground rocks, wireless communication equipment is difficult to use. This creates a significant communication barrier, reducing rescue efficiency significantly. Participant 2 has worked in firefighting for over 30 years. He has witnessed the development of the entire communication equipment of China's emergency rescue team, from the first radio station to today's satellite communication. He stated that with technological advancements, new communication equipment has provided them with better solutions. However, in disaster relief operations, response teams continue to face various communication barriers. In large-scale natural disaster rescue, for example, the use of various communication devices frequently necessitates diverting their own attention to ensure contact with team members and other teams. Participant 6 reported that in the rescue of high-rise building fires, communicating with teammates and confirming the position of teammates was a more difficult thing.

Workload

Emergency rescue entails a heavy workload, and this is frequently a major issue that plagues first-line rescue workers. They are confronted with the load rescue environment and noise pollution. At the same time, they must ensure the safety of the rescued personnel while also protecting themselves and their teammates. Also, the cognitive workload is also a problem for rescue participants. Especially in the rescue of fire and building ruins, the cognitive workload has increased significantly. In a complex disaster environment, the complex site conditions will increase the cognitive load of rescue workers. Participant 8 has a long history of firefighting experience. He believes that firefighters face a massive workload when it comes to fire rescue. Firefighters must wear heavy fireproof clothing while searching for trapped people through the fire and smoke. Participant 8 also mentioned in the interview that as a professional firefighter, he should not only pay attention to the trapped people, but also pay attention to the development of the fire and the destruction of buildings. This is a huge challenge for firefighters' cognitive workload. When discussing the workload in earthquake relief, the second participant stated that the terrain often changes dramatically in large earthquakes. These changes require them to use more energy, as well as using a variety of equipment to locate trapped people in buildings that could collapse at any time. These duties significantly increase their workload. At the same time, many participants reported that in the large-scale disaster scene, such as earthquake rescue, because of the damage of the earthquake to the landform, the rescue personnel not only need to complete the rescue of the trapped, but also need to identify the risk through careful observation of the scene. This has also greatly increased their cognitive workload.

Navigation

Navigation is a significant human factors issue in emergency rescue. From various perspectives, all nine participants described the difficulties they had encountered while navigating. Participants 1, 5, and 8 are all emergency management personnel from China, and have all assisted in the rescue of numerous production safety accidents. These navigation issues were most acutely experienced in chemical plant accident rescues and mine navigation in coal mine safety accidents. Due to the presence of various pipelines and equipment in the factory, rescue workers face complex search routes. Due to the complexity of underground tunnels and route changes caused by various collapses, mine rescue workers frequently face difficult navigation routes. Participants 2 and 8 have previously helped with earthquake relief. They stated that due to the topographic changes caused by the earthquake, it was difficult for them to locate their true location on the map. The navigation problem in fire rescue was mentioned by five participants from the fire department. They believed that the fire and smoke had hampered their navigation.

Stress

Stress is another major human factor that afflicts emergency staff. High-intensity work in disaster relief has put rescue workers under a lot of strain (related to *Workload*). At the same time, participants reported on their stress levels following the disaster. This is because in disaster rescue, sometimes they will face the situation that not all survivors can be rescued, and the response teams have to make a difficult choice about who to rescue. This not only causes them to face great pressure during the process of disaster relief, but also after the rescue action, they sometimes fall into self-blame. This has also affected their mental recovery after the disaster.

VR and AR application

Following the explanation and presentation of VR and AR applications, each participant was enthusiastic about the use of these technologies in emergency rescue. However, Participant 2 mentioned that the heat from a fire would damage the VR/AR helmet or glasses if used during firefighting. He also brought up concerns about the battery life; rescuers are frequently faced with long-term continuous work in real-world scenarios, and he questioned whether VR/AR equipment could support long-term work. A communication issue with VR/AR equipment was raised by participant number four. The communication infrastructure is frequently severely damaged in natural disasters such as earthquakes and floods. His concern was how to use the equipment under such communication conditions. Participant 6 inquired about 3D vertigo. He mentioned that he had used VR devices before, but they made him dizzy, and was concerned about this if the technology was adopted in emergency response efforts. However, he was very interested in the technologies and hoped to learn more about the use of VR and AR in emergency management. Participant 1 reported being impressed by a VR training facility for earthquake escape in a civic disaster reduction education centre. He believes that if such equipment can be used in rescue work to solve problems such as teamwork, navigation, communication, and so on, it will be a huge step forward in the rescue mode.

Findings

The results showed that good teamwork is the most important factor influencing the success of emergency rescue missions. However, communication, navigation, and workload are also important factors in team cooperation. These factors interact with one another, resulting in an emergency rescue that is inherently complex. Interviewees explained that this is partly due to the unpredictable nature of disasters: emergency managements team composition and activities are not consistent across all events. Thus, cooperation and communication across teams are essential factors determining the efficiency and success rate of rescue activity as the teams respond to each individual event.

Interviewees reported that VR and AR has the potential to alleviate some of the human factors challenges in emergency rescue and team cooperation by supporting visualisation of data from existing technologies (e.g. sensors and scanners). Simultaneously, the integration of VR and AR-based technologies with communication, navigation, and various other rescue equipment may help the emergency team improve its work efficiency.

Discussion and Conclusion

This paper reports on in-depth conversation with interviewees about the human factors in emergency rescue that I discovered. This confirmed that teamwork, communication, navigation, and workload issues discovered in the literature are, in fact, the main issues confronting the emergency rescue team.

This paper presents a more detailed depiction of the human factors concerns that are associated with emergency management. When dealing with a wide variety of crises, emergency managers face a number of challenges related to human factors. Working together effectively is one of the most difficult human factors challenges that emergency management teams must overcome. Communication, workload, stress levels, and perceptions are all examples of human factors that play a role in determining the efficiency with which emergency management teams collaborate. This is because of the specific nature of emergencies; teamwork in the context of emergency management teams entails more than just cooperation between the individuals that make up the team. As a result of the complexity of emergency situations, it is necessary for emergency management teams from various localities, regions, and even countries to collaborate with one another. Cooperation between different emergency management teams is therefore another significant challenge that they face.

Emergency management teams can combine virtual environments with real-world disaster scenes using augmented reality technology. This will help frontline emergency responders reduce the amount of work they have to do while also improving the communication and collaboration within their teams. The virtual reality technology can bring a more realistic representation of the disaster scene to the command centre. It is anticipated that emergency team commanders will be better able to direct their teams and make decisions if virtual reality technology is utilised at the scene of a real disaster. When the benefits of augmented and virtual reality are combined, emergency management teams will have access to a new method of operation that will assist them in better coping with the human factors that are involved in responding to emergency situations. However, human factors are an essential foundation to any technological development work in this area.

References

EFRA. (2021). Flooding fourth report of session 2019–21. https://committees.parliament.uk/publications/4601/documents/46603/default/

Fischer III, H. W. (1996). What emergency management officials should know to enhance mitigation and effective disaster response. Journal of Contingencies and Crisis Management, 4(4), 208–217. https://doi.org/https://doi.org/10.1111/j.1468-5973.1996.tb00095.x

Frye, L. M., & Wearing, A. J. (2016). A model of metacognition for bushfire fighters. Cognition, Technology & Work, 18(3), 613–619.

Goldiez, B. F., Ahmad, A. M., & Hancock, P. A. (2007). Effects of augmented reality display settings on human wayfinding performance. IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), 37(5), 839–845. https://doi.org/10.1109/tsmcc.2007.900665

GOV.UK. (2022). Flooding and coastal change: Detailed information. https://www.gov.uk/topic/environmental-management/flooding-coastal-change

Hayes, P. (2017). The impact of team member familiarity on performance: Ad hoc and pre-formed emergency service teams. Human Factors Challenges in Emergency Management.

Johnson, C. (2017). Expert decision making and the use of worst case scenario thinking. Human Factors Challenges in Emergency Management.

Kapucu, N., & Garayev, V. (2011). Collaborative decision-making in emergency and disaster management. International Journal of Public Administration, 34(6), 366–375.

Kapucu, N., & Hu, Q. (2016). Understanding multiplexity of collaborative emergency management networks. The American Review of Public Administration, 46(4), 399–417.

Lisa M. Frye, A. J. W. (2017). What were they thinking? a model of metacognition for bushfire fighters. Human Factors Challenges in Emergency Management.

McLennan, J., Strickland, R., Omodei, M., & Suss, J. (2017). Stress and wildland firefighter safety-related decisions and actions. Human Factors Challenges in Emergency Management. Ashgate Publishing, Ltd., 19-31.
Murphy, B. L. (2007). Locating social capital in resilient community-level emergency management. Natural Hazards, 41(2), 297–315. https://doi.org/10.1007/s11069-006-9037-6
Owen, C. (2014). Human factors challenges in emergency management: Enhancing individual and team performance in fire and emergency services. Ashgate Publishing, Ltd.
Owen, C. (2017). Leadership, communication and teamwork in emergency management. Human Factors Challenges in Emergency Management. Ashgate Publishing, Ltd., 125-144.
Park, S., Park, S., Park, L., Park, S., Lee, S., Lee, T., Lee, S., Jang, H., Kim, S., Chang, H., & et al. (2018). Design and implementation of a smart iot based building and town disaster management system in smart city infrastructure. Applied Sciences, 8(11), 2239. https://doi.org/10.3390/app8112239
Wilson, J., & Oyola-Yemaiel, A. (2001). The evolution of emergency management and the advancement towards a profession in the united states and florida. Safety Science, 39(1–2), 117–131. https://doi.org/10.1016/s0925-7535(01)00031-5

Assumption-Based Leading Indicators to Monitor Healthcare System Drift Towards Failures

Elizabeth Wilson-Lewis & Gyuchan Thomas Jun

Loughborough University, UK

SUMMARY

This study explores how the assumption-based leading indicator approach can be applied to monitor organisational drift of healthcare systems towards failures based on the Mid Staffordshire NHS hospital failure case.

KEYWORDS

Assumption, leading indicators, healthcare and practical drift

Introduction

Today, healthcare systems are becoming increasingly more complex due to highly interconnected elements including institutions, people, processes, and resources (Bashford, et., 2018). Adverse events in healthcare are the fourteenth leading cause of the global disease burden (Slawomirski et al, 2017). These healthcare adverse events significantly represent the safety management of an organisation, values, attitudes, and behaviour patterns. Organisations tend to gradually erode without the organisation being aware of this drift, which is often an important precursor to organisational accidents. Leveson (2017) argues that most adverse incidents can be controlled and detected through risk monitoring systems such as leading indicator monitoring programs. Zwetsloot et al., (2020) found that proactive leading indicators are intended not only to better direct and control the safety, health, and wellbeing of the system, but also to support the development of prevention culture. Wormaes (2015) claimed that highly complex socio-technical systems like healthcare which are dominated by human actions and interaction, will require a new leading indicator framework. The concept of assumption-based leading indicators implies that a warning sign can be used in monitoring a process to detect when an assumption is broken or dangerously weak or when the validity of an assumption is changing (Leveson, 2014). Assumptions are defined as a 'belief or feeling that something will happen, although there is no tested proof.' Assumptions play an important role in developing a safety control structure and assigns responsibilities for the system. In order to avoid any system failures, it is important to make sure that the assumption-based plans are not vulnerable to violations or unplanned changes throughout the organisational operations. In addition, operational systems should always be monitored to ensure that the system is operated and maintained in a manner assumed by the designers. The aim of this study is to explore how the assumption-based leading indicator approach can be applied to healthcare system safety management.

Method

We used the Mid Staffordshire NHS hospital failure case (Francis 2013 public inquiry report) which is an example of a large-scale healthcare system failure in the UK (400 and 1,200 patients died because of poor care over the 50 months between Jan 2005 and March 2009). We created a hierarchical control structure model (STAMP-based) of the healthcare system relevant to the failure; STAMP explains that accidents are a result from inadequate enforcement of system safety constraints in design, development, and operations (Leveson et al., 2015). A documentary analysis of the Mid Staffordshire public inquiry report (executive summary) was conducted using NVivo software to identify incorrect assumptions that existed or became incorrect over time in terms of feedback and control actions. Furthermore, an ultimate question put forward was what assumption-based leading indicators can be put in place to monitor healthcare system drift towards failures.

Results and Conclusions

Table 1 shows the results of the analysis. The findings show that various inadequate visible behaviours or actions of stakeholders at the multiple levels are based on broken or invalid assumptions about feedback they are receiving or control actions they assume that other actors might take. In general, actors naively believed that limited feedback they get are accurate indicators of what is going on or sometimes selectively accepted certain assuring feedback while ignoring more concerning feedback. It seems that various cognitive bias, e.g. belief bias, availability heuristics, confirmation bias, influenced the broken or invalid assumptions people made behind their inadequate behaviours and actions. The findings have important implications for developing assumption-based planning to ensure that the assumption-based plans are not vulnerable to violations or unplanned changes throughout the organisational operations (Leveson, 2014). The assumption-based planning can complement traditional risk/hazard assessment approaches by ensuring that the models and assumptions used during the initial decision making and planning were appropriate and support the organisation to be resilient towards vulnerable assumptions.

Table 1: Leading indicators and underlying assumptions – Mid Staffordshire NHS Failures

System levels	Leading indicators (behaviour/actions)	Broken or invalid assumptions behind leading indicators (Feedback/Control actions)
NHS regional oversight	Relaxation of safeguards & controls	**Feedback** ▪ Intelligence from others would be accurate ▪ Adequate patient safety monitoring systems would be in place ▪ Transfer of operational information would be clear/adequate **Control action** ▪ Other actors would be proactive in managing high risks
Health Board Trust	Too much emphasis on the Foundation Trust application	**Feedback** ▪ It would be okay to accept the reported false assurances, (e.g., relying on mortality ratios rather than the reported system violations) **Control action** ▪ Releasing the financial pressure by becoming the Foundation Trust would be most important
	Allocation of mismatched tasks to healthcare workers who were not adequately trained or qualified	**Control action** ▪ It would be okay not to plan for future uncertainties ▪ Staff would be able to cope with unplanned changes, unintended goals, & inadequate system designs ▪ Reported complaints from patients, families, and community would be adequately addressed or resolved in a timely manner
Divisional Management	Bullying care team members to keep their heads down	**Feedback** ▪ There would be continuously managed adequate risk planning & monitoring framework in place **Control action** ▪ Relying on past experiences while not accepting or being prepared for evolving system changes & patient needs

Hospital Care Team	Failed attempt to report patient safety concerns	**Feedback** ▪ There would be accurate feedback channels of pertinent information during unplanned changes that would be timely, communicated, and transferred between all operational levels both internal & external

References

Bashford, T., Dean, J., Clarkson, J Ward, and. J., Komashie, A. (2018). A systems approach to healthcare: from thinking to practice. Future Healthcare Journal, 5(3), pp.151–155.

Leveson, N. (2014). A Systems approach to risk management through leading safety indicators. Reliability Engineering & System Safety, 136, p 17-34.

Leveson, N., Daouk, M., Dulac, N. and Marais, K. (2015). Applying STAMP in Accident Analysis. [online]. Available at: https://shemesh.larc.nasa.gov/iria03/p13-leveson.pdf.

Leveson, N. (2017). Engineering a safer world: systems thinking applied to safety. Cambridge, Massachusetts: The MIT Press.

Slawomirski, L., Auraaen, A. and Klazinga, N. (2017). The economics of patient safety strengthening a value-based approach to reducing patient harm at national level. OECD

Wormnæs, M. (2015). Leading indicators for real-time monitoring of risk in healthcare organizations. Master's thesis at the University of Stavanger.

Zwetsloot, G., Leka, S., Kines, P. and Jain, A. (2020). Vision zero: Developing proactive leading indicators for safety, health, and wellbeing at work. Safety Science, v130.

Using Systems Thinking to Identify Risks in Telephone Triage: MEAD Study Findings

Jill Poots[1,2], Jim Morgan[1], Matteo Curcuruto[1], Stephen Elliott[2], & Andrew Catto[2]

[1] Leeds Beckett University, United Kingdom [2] Integrated Care 24 (IC24), United Kingdom

SUMMARY

This paper presents findings from a modified Macroergonomic Analysis and Design (MEAD) study aiming to identify system components and risks in a telephone triage system. Themes identified included: '*accessibility and availability*'; '*risks on the part of the telephone triage professional*'; '*risks posed by callers*'*;* and '*barriers to safety incidents*'.

KEYWORDS

Telephone triage, primary care, macroergonomics

Introduction

Telephone triage use is rising, due to its convenience, advantages for disease control, and high patient satisfaction. Whilst considered to be predominantly safe, telephone triage services have been implicated in harm to, and death of, patients (Rees et al., 2017). Despite these concerns, there has been little attempt to study telephone triage work through a human factors lens, to improve patient safety. To address this research gap, this study used a modified Macroergonomic Analysis and Design (MEAD) approach (Murphy et al., 2018) to identify system components, their interactions and contributions to risk in a telephone triage system. A secondary aim was to assess the suitability of the MEAD framework for exploring factors affecting safety in complex sociotechnical systems.

Methods

Murphy et al.'s modified MEAD framework was used to understand system components, their interactions and subsequent risks for patients, via the following steps: initial system scan; key informant interviews (n =25) using the critical incident technique (CIT); analysis using mapping; and validation of findings. To understand interacting components, while mitigating researcher bias, Leximancer software was used to analyse interview findings. These findings were then validated through reflexive thematic analysis. To visualise deviations from intended system use and design (i.e. variances), and depict interactions between system components, a variance matrix (Kleiner, 2002) was also constructed. Variances were developed by comparing document scans and interview findings, before being presented to subject matter experts, with the aim of validating findings and ascertaining future research priorities.

Results

System Mapping

The system map was developed iteratively using document scans, discussions with subject matter experts, and interview findings, and subsequently illustrated using LucidChart. It revealed there are a number of 'work systems' encountered in any one triage call, many of which involve increased patient-professional collaboration and human-computer interaction. For example, at several points of the system, communication between a patient and professional is mediated by technology.

Leximancer analysis

Leximancer indicated '*time*' as one of the most salient risk factors, across various points of the system (for example, lengthy delays accessing the service, or receiving clinical advice from colleagues). Concepts relating to the external environment such as '*ambulance*' and '*COVID*' were also frequently mentioned, suggesting the telephone triage service does not operate in isolation, but relies on other parts of the health system to work well. Technology-related concepts highlighted the important role of computer decision support systems in mediating communication between patients and professionals.

Validation and construction of the variance matrix

Secondary analyses revealed similar trends to Leximancer, and yielded different data extracts, extending and validating the findings. Perceived risks according to advisors resulted in themes such as: '*risks on the part of the telephone triage professional*', '*risks posed by callers*', '*risks due to accessibility and availability*' and '*barriers to safety incidents*'. For example, '*time*' was again mentioned, with respect to the potential for patients' health to deteriorate if they have to wait at multiple points in the system. Variances identified included conflict between roles due to key performance indicators, and an under-appreciation of the role of patients in their care. These were validated using recent focus group with sixteen representatives from within the organisation using a Likert-scale. Agreement was high for all proposed variances, risks and the proposed system map.

Discussion

MEAD proved a useful framework for mapping the system and identifying risks and variances in a service not previously investigated using human factors tools. It positioned this telephone triage system within the context of the external environment, suggesting availability and public health issues may impact the likelihood of safety incidents in telephone triage systems. Risks unique to telephone triage were outlined, such as the role of the patient in reporting their symptoms accurately. It is anticipated the validation of the variance matrix will be useful in identifying research priorities and serve as a foundation for more specific human factors research. Although participants were recruited from multiple contact centres and levels of the host organisation, it would be useful to extend this research to other providers of telephone triage including for-profit enterprises and GP surgeries. A future modified Delphi study aiming to identify contributory factors specific to telephone triage systems will include experts from a broader sample of organisations.

References

Kleiner, B. M. (2002). Macroergonomic Analysis and Design (MEAD) of Work System Processes. Human Factors and Ergonomics Society 46th Annual Meeting,

Murphy, L. A., Robertson, M. M., Huang, Y.-h., Jeffries, S., & Dainoff, M. J. (2018). A sociotechnical systems approach to enhance safety climate in the trucking industry: Development of a methodology. *Applied ergonomics, 66*, 82-88.

Rees, P., Edwards, A., Powell, C., Hibbert, P., Williams, H., Makeham, M., Carter, B., Luff, D., Parry, G., Avery, A., Sheikh, A., Donaldson, L., & Carson-Stevens, A. (2017, Jan). Patient Safety Incidents Involving Sick Children in Primary Care in England and Wales: A Mixed Methods Analysis. *PLoS Med, 14*(1), e1002217. https://doi.org/10.1371/journal.pmed.1002217

Key User Factors in a New Style Two-Person Train Cab

David Hitchcock[1], Kimberley Harding[2] & Penny Gazard[3]

[1]David Hitchcock Limited, [2]University of Birmingham, [3]Transport for Wales

ABSTRACT

The introduction of a new train to Transport for Wales (TfW) was initially faced with reluctance and resistance from key stakeholders. Consistent with previous train driver projects, TfW elected to adopt an ergonomics approach to identify, clarify and resolve issues of acceptance, safety and performance. The reasons for this were to address cab design issues, and to inform the design and delivery of a training programme intended to facilitate a cultural shift in drivers' ways of working and those who support them. The ergonomics investigative work embraced consultations, literature review, measurements and observations of cab users, and was conducted alongside, and in harmony with, engineering reviews. This integrated approach led to a successful resolution to the concerns of stakeholders, resulting in accepted changes to working practices and deployment of the train.

KEYWORDS

Cab, Driver, Instructor, Train, Triangulation

Introduction

The introduction of new rolling stock as part of a major transformation of Transport for Wales (TfW), the Stadler FLIRT fleet, with its progressive cab design, required user acceptance. As a matter of prudency, founded on a legacy of considering human factors, the introduction of this new train model prompted an assessment of factors that could influence the acceptability, safety and performance of train drivers and those who instruct and monitor them.

Figure 1. The Stadler FLIRT Cab.

The successful design and provision of a training programme to enable experienced and new drivers to operate the new train was vital his work. This programme necessitated engagement with various stakeholders because of the 'TfW Social Partnership' agreement of using collaborative methods to bring about a cultural shift in drivers' ways of working. Key stakeholders were Trade Unions and Health and Safety Representatives.

It was clear in the early stages of the training design project that there was resistance to the new cab design, principally because of two key design features new to the cab users:

- The driver seat was centralised in the cab, not positioned on the left-hand side as the drivers were used to.
- The second-person (fold-down) seat was positioned on the left-side of the driver rather than the more familiar, right-hand side

The design of the new cab was considered unacceptable by Trade Union partners, and without their buy-in, bringing about the transformation of TfW's network, through the introduction of new rolling stock, would be nigh-on impossible. Furthermore, the introduction of considerable new technology in the cab, added to the seating and cab layout concerns which resulted in many drivers being apprehensive and unwilling to undertake training.

The Stadler FLIRT trains with this cab design had already successfully entered service with another UK TOC, although the evidence for this deployment available to TfW was largely anecdotal, and certainly did not address the issues highlighted by TfW's Social Partners. It was, therefore, agreed that an evidence-based approach was paramount. This further validated the application of triangulation.

In this instance, the data and information to be used in the triangulation were collected through, literature review, observation of users, consultation with users, and measurement of cab layout. These were all within the context of stakeholder engagement, including liaison with personnel from the project team, engineering, operations, safety and unions.

The programme involved three main stages:

Stage 1: Assessment of acceptance factors for the primary and secondary users. This was conducted on a static train in the depot while the users simulated the tasks they would perform in reality. This was followed by consideration of the same factors with the test train on the move (without public passengers).

Stage 2: Several factors identified worthy of further assessment were subsequently investigated.

Stage 3: Following a review of the findings, it was decided by the project team to more closely review the potential impact of distraction to drivers resulting from the cab design and its use.

The Approach

Consistent with previous train driver projects with TfW, this work adopted the aforementioned triangulated approach (Denzin, 2007). Triangulation is an accepted method used in situations where subjectivity plays a significant role. By considering the same factors from different perspectives, triangulation seeks to clarify agreements, yielding greater confidence in findings. Triangulation was considered especially important in this project, because the real-world constraints of the operating company (TfW) and the circumstances of the work (e.g. constraints on availability of trains and track time), meant that only a limited number of trains, instructors and drivers would be involved in work of this nature.

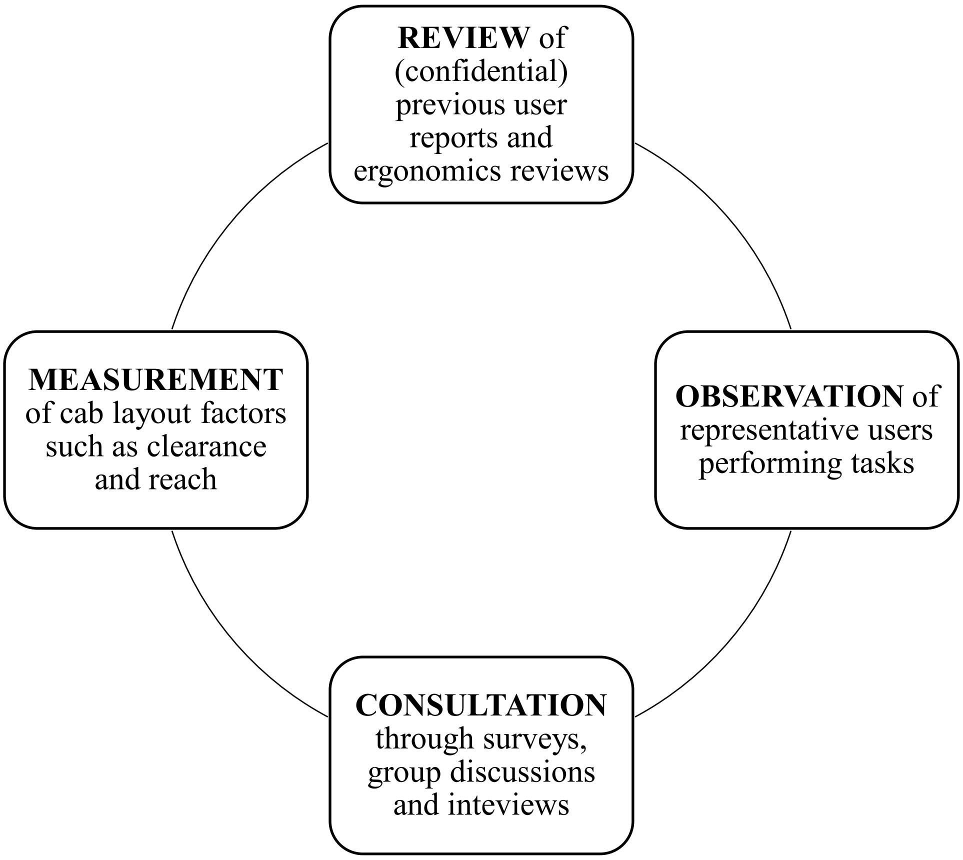


Figure 2. Data And Information Sources Used for Triangulation in This Work.

Stage 1: Assessment of Acceptance Factors

The cab acceptance factors considered included (in alphabetical order); access and egress, clearances and reaches, comfort of the seats, ease of setting up the seats, position and clearance for lower limbs, sense of security/stability, viewing provision, and working space and position.

User participation was core to the ergonomics work (Hitchcock et al. 2016), so with the intention of providing a realistic and pragmatic representation of users in the 'acceptability trials' (observation and consultation) the participants comprised (self-reported):

- Age category range: (25-34) to (55 or over).
- Gender: 3 female,9 male.
- Stature range: between 5%ile and 85%ile UK working age adult (Open Ergonomics, 2020).
- Weight range: between 5%ile and 85%ile UK working age adult.

The trials took place, either on a static train in the depot or on a moving train (without passengers). In both cases, participants were asked to simulate and/or actually perform the full range of tasks they might normally be engaged in (e.g. *"…look at what you might need to look at, reach for what you might need to reach, press what you might need to press…"*).

The participants were observed in action and at the end of each trial, all users completed a survey (feedback form), which chiefly asked them to rate the factors on a 7-point, bi-polar scale. For illustration, Figure 3 presents two summary charts of these ratings.

'Driver Seat' refers to the central seat of the primary user (the Train Driver). 'Second Seat' refers to the other seat in the cab, used by the secondary users, typically instructors (the user type given focus in this work).

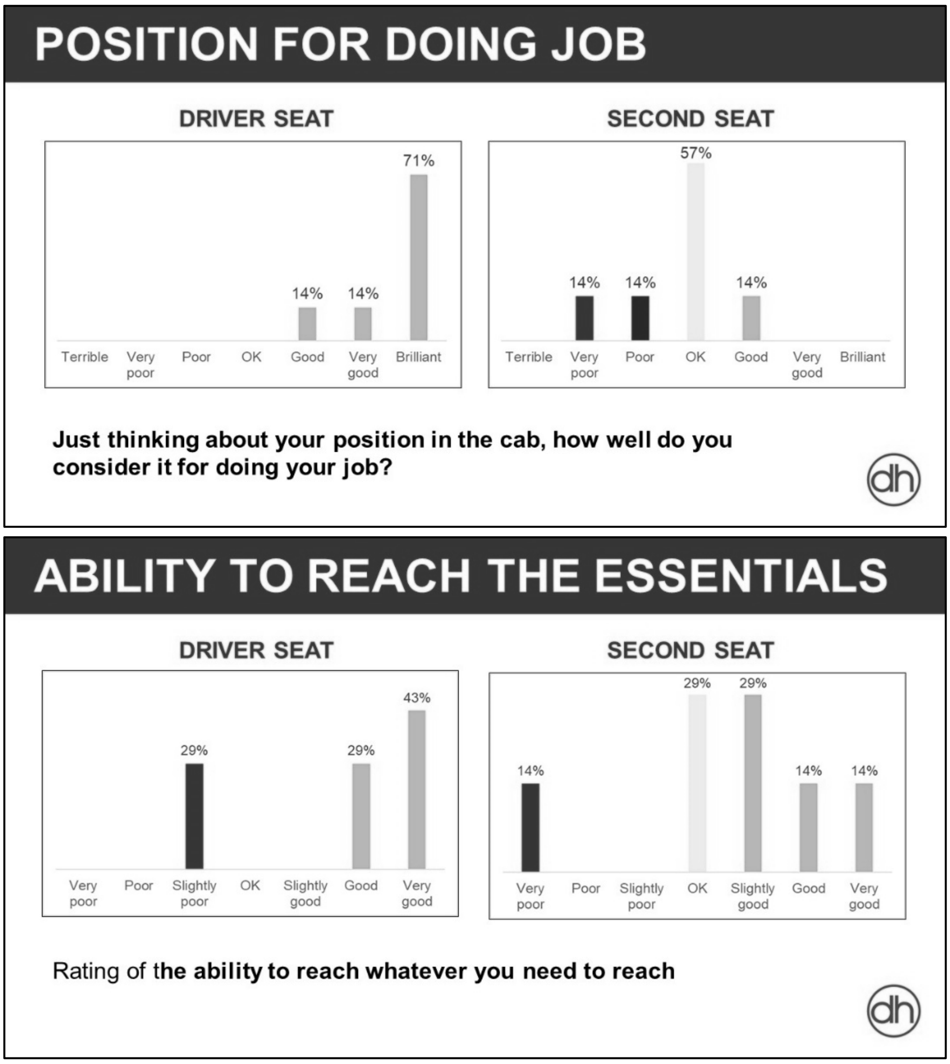


Figure 3. Illustrative Charts of The Acceptability Trials Survey Feedback.

Headline findings of this first stage of the ergonomics input were:

- Overall, 70% of the factors considered were rated 'OK' or 'Better' in the user feedback survey.
- The primary user driver seat received a very good assessment for its design, comfort and position. This, alongside other clearance and reach factors were in agreement with the anthropometric assessment using the Railway Safety and Standards Board MAT tool.
- The secondary user (e.g. instructor) seat received a less favourable acceptability assessment, principally due to its close proximity to the driver seat. This rendered it important to risk assess access and egress, review position with respect to track signal viewing and monitor comfort factors.

Stage 2: Investigation of Emerging Issues

Although the observations were not analysed in detail (there was no need, given the nature of the factors being considered) it was apparent that they were reflected in the consultation feedback; and were not unexpected given the review of the confidential literature sources. Consequently, three issues emerged as warranting further investigation. All three were considered safety/performance critical for both the primary and secondary users. Importantly their attention demonstrated a growing commitment between the stakeholders to address issues of concern. This unquestionably led to increased iteration of ergonomics application and closer engagement with Social Partners.

Proximity of the Second Seat to the Driver Seat

Even though at the early stage of the ergonomics work there was insufficient evidence to conclude that the two seats were actually too close together to be acceptable, the opportunity was taken to conduct an engineering review of the feasibility of moving or modifying the second person's seat. The review determined that space restrictions made such changes unrealistic:

- On the left-hand side, there was very limited space due to the Power Converter.
- On the right-hand side, there was very limited space due to the main Pneumatic and Electric Control Panels.
- Adding a small perch seat to the right-side of the driver would unacceptably reduce the width of the safety egress throughway to the external door.

Cab Access & Egress

Access and egress issues were raised by the participants in the trials as being potentially problematic, so these were investigated, and the risk assessed through a simulation exercise conducted on a static Stadler FLIRT train using a powered vehicle ensuring the driver's seat had full air supply so that it represented real-in-use conditions. 5 drivers ranging in height and body mass were shown the driver's seat features (the seat has 8 adjustable features) ensuring they were able to set up the seat to their preferred, natural, driving position. Each participant used the driver's seat and in turn occupied the second seat.

It was found that - depending on the driver seat setting - the weight bearing fold-down/flip-up fixed position, second seat was restricted – both for deployment and return. Consequently, the participant drivers were requested to problem solve. Organically, they worked through various methods, before a best practice solution was established. This was captured in a risk assessment which led to a best practice supplementary user guidance document, issued during, and supported in the new fleet training programme materials.

Instructor (Secondary User) Viewing

Although, unlike for drivers, there are no specific standards with which to comply, the ability for an instructor to see the likes of trackside signals is a key aspect of effective training.

To check if this was achievable in the cab, given the centralised driver position, an exercise was conducted in-house in which two people of the design expectation extreme percentiles (5^{th} and 95^{th}) were used to view a target red signal post. It was considered helpful by the project that the exercise followed the same approach used to assess the compliance with industry standards of driver signal sighting.

The exercise was conducted on a static Stadler FLIRT train and recorded viewing measurements from three cab positions:

- Seated in the second seat positioned to the left of the driver.
- Seated on a corresponding movable stool, positioned to the right of the driver.
- Stood in corresponding position on either side of the driver.

The exercise found that the viewing experiences of both percentiles were different during standing and sitting positions:

- The 5^{th} percentile could view more from the standing position rather than sitting.
- The, 95^{th} percentile was able to view more from the sitting position than standing.
- The optimal views for both percentiles were achieved by positioning themselves to the adjacent side of the driver's seat when viewing targets to the opposite side to where they were stood.

Consideration was, therefore, given to, what might be described as, changes in posture which occurred naturally, such as lowering or lifting the head, or changing standing position from the left to the right of the driver and vice-versa. From a performance perspective, these changes were evidently beneficial. Furthermore, such changes were supported by the literature (Buckley et al., 2015, Black et al., 2022) which could be surmised as: 'the best posture is the next posture'.

Stage 3: Consideration of Distraction Concerns

The final 'people aspect' to be reviewed before the fleet could be released into action was to consider the impact on the driver of the secondary cab user moving around the cab or adopting different positions within the cab in order to best perform their duties. In the same vein, the work needed to monitor how the secondary user might be impacted –were they concerned that their movement around, and proximity to, the driver adversely affect the training?

Therefore, to maximise the potential number of user engagements it was decided to conduct this work as part of the programme of training of drivers to use the new cab, and to receive feedback from both the drivers and instructors. This feedback was consolidated with observations to triangulate feedback information. In addition, new trainees (inexperienced drivers) were also observed and consulted but using the Stadler FLIRT simulator rather than an actual live train on tracks. Within the timeframe of the programme, 15 drivers and 6 instructors were observed. In both cases, many participants submitted multiple feedback forms because they received or gave multiple training sessions. A greater sample of drivers were consulted through group discussions and individual interviews.

Table 1. Headline findings of the Distraction Consideration Exercise

Driver Survey Feedback of Occurrences of Distraction	Instructor Survey Feedback of Perception of Potentially Causing Distraction	Comments
When instructor was sat on second seat, 70% of responses were 'Very Small Distraction' or less.	When instructor was sat on second seat, 75% of responses were 'Very Small Distraction' or less.	The survey findings reflected the information gathered through the other sources of triangulation. The instructors were more sensitive to the possibility of causing distraction than was actually realised; this was anecdotally attributed to their professional practice.
When instructor was stood to left of driver, 91% of responses were 'Very Small Distraction' or less.	When instructor was stood to left of driver, 90% of responses were 'Very Small Distraction' or less.	
When instructor was stood behind driver, 87% of responses were 'Very Small Distraction' or less.	When instructor was stood behind driver, 64% of responses were 'Very Small Distraction' or less.	
When instructor was stood to right of driver, 91% of responses were 'Very Small Distraction' or less.	When instructor was stood to right of driver, 89% of responses were 'Very Small Distraction' or less.	
When instructor was moving around cab, 92% of responses were 'Very Small Distraction' or less.	When instructor was moving around cab, 71% of responses were 'Very Small Distraction' or less.	

These findings supported the preceding work of the benefits to instructor viewing performance and making changes to their posture and movement. Indeed, the instructors reported that working in different positions around the cab enabled them to perform their tasks better and perhaps even reduced the distraction; for example, by being able to see the speedometer rather than repeatedly asking the driver to tell them the speed. In a complementary manner, during group discussions and interviews, drivers indicated that such practice was preferable as it seemed easier to have different controls and displays interactively explained and demonstrated.

Key Conclusions of the Work

Because of the potentially significant safety impact, a primary concern of the acceptance work was the possible cab access and egress limitation – despite receiving generally good feedback. This was further investigated by TfW and a protocol established for efficient and safe access and egress.

Impressive initial feedback from the drivers (for the cab in general and their seat), remained consistent throughout the different ergonomics investigations.

The second seat, initially, received less positive feedback, essentially because of its lack of adjustability, relative 'lack of luxury' compared to the driver seat and its close proximity to the driver seat. Nevertheless, the ratings were reasonable; and furthermore, as the ergonomics investigations progressed, less and less criticisms were aired. Indeed, the feedback moved toward a recognition that it may well represent the best second seat across all of TfW stock.

Instructors reported varying degrees of comfort in the initial work, but this similarly changed as they became used to the benefits of working in different positions. These included being able to change posture and being able to position themselves for instruction for better viewing inside and outside of the cab and not needing to enter driver personal space in order to highlight controls and displays.

Initially instructors appeared to be more concerned that their movements around the cab to change positions might be causing distraction to the driver than the drivers were actually reporting. However, by the conclusion of the ergonomics investigation, it was evident that both user types considered things to be (quote): "*business as usual*" and that distraction from instructors did not present a problem.

Lessons Learned

Perhaps most notably, the experience gained from this work will influence future fleet projects, particularly the use of a thorough, triangulated ergonomics approach.

The results from this work suggest that a new fleet training programme has facilitated a culture shift, which meets the operational requirements to introduce a train into TfW.

In the rail sector, non-technical skills are common in training and development, particularly in operational safety critical roles; this work has highlighted the essential role of ergonomics throughout a project, especially one which focuses on a new working environment.

Throughout the work, high levels of communication between a wide group of stakeholders was maintained and anecdotally was considered beneficial in the adoption ('buy-in') process. For example, presentation sessions were hosted for all TfW Operational Managers so they could understand the rationale behind competency management system enhancement which was developed as training requirements in response to the findings of this work, including:

- Best behavioural practices when two users are in the cab.
- How to set up the seats and their footrest, considering the associated risks.
- Safe access and egress.

What Next?

At the final presentation of this work, it was agreed that the training programme could continue for another three months and that feedback surveys would also continue to be collated to provide more data from a wider pool of users. This could be considered as further recognition of the benefits of ongoing ergonomics contribution throughout a project of this nature.

References

Black, N.L., Tremblay, M., Ranaivosoa, F. (2022). Different Sit:Stand Time Ratios Within A 30-Minute Cycle Change Perceptions Related To Musculoskeletal Disorders. Applied Ergonomics, Vol. 99.

Buckley, J.P., Hedge, A., Yates, T., Copeland, R.J., Loosemore, M., Hamer, M., Bradley, G. and Dunstan, D.W. (2015). The Sedentary Office: An Expert Statement On The Growing Case For Change Towards Better Health And Productivity. British Journal Of Sports Medicine, 49(21), pp.1357-1362.

Denzin, N.K. (2007). Triangulation. The Blackwell Encyclopedia Of Sociology.

Hitchcock, D., & Campbell, C. (2016) Is It Really Worth Asking The Users? A Case Study from the Rail Industry. Contemporary Ergonomics and Human Factors 2016. Eds. Rebecca Charles and John Wilkinson. CIEHF.

Kuniavsky, M. (2003). Observing The User Experience: A Practitioner's Guide To User Research. Elsevier.

Open Ergonomics (2020), PeopleSize 2020 Professional.

Railway Safety and Standards Board (2015), Musculoskeletal disorder risk Assessment for Train drivers (MAT) tool (version 2.0).

A Quantitative Approach to Determining Inclusive Design Features Within UK Railway Depots

Kimberley Harding[1], David Hitchcock[2]

[1]Birmingham Centre for Railway Research and Education, University of Birmingham, UK, [2]David Hitchcock Ltd, UK

ABSTRACT

Inclusive design is fast becoming a talking point for engineers, ergonomists, and designers. With holistic views on accessibility and inclusion in railways at the forefront of modern-day design and culture, it is a novel concept which could drastically change how we live and work. However, despite common interest in pursuing inclusive design strategies for railway passengers, there are significant gaps in inclusive thinking for railway workers – namely those who work in depots and trackside. Despite reforming modern system safety techniques, 2022 saw a significant increase in workforce injuries and little to no assistance in returning to work post-injury (Johnson, 2023). Limited accessibility and manoeuvrability in railway depots and trackside sees a workforce of injured staff unable to properly return to their original place of work and having to re-train in a less demanding sector.

In response to this, this paper presents findings and methodologies for quantifying the inclusiveness of depot design from a user-centred approach. The data captured emulates how staff experience working in train maintenance and what aspects could be improved to encourage inclusivity in the workplace whilst enabling staff to work to their highest potential. It is hoped that this research could reduce bias in quantifying inclusive design elements in depots, framing a new scope for what is deemed 'accessible' or 'non-accessible' to make railway depots a better working environment for all.

KEYWORDS

Train 1, Depot 2, Inclusivity 3

Introduction

Inclusive design is an element that wholly encompasses all users irrespective of 'age, gender and disability' (Inclusive design hub, 2021). The user-centred design approach allows ergonomists and design engineers to align their work with user requirements through all design stages, bettering work at every stage in the design process.

In a climate where 'over 1 billion people' (World Health Organisation, 2020b) have a long-term disability and 'almost everyone will temporarily or permanently experience disability at some point in their life' (World Health Organisation, 2020a), inclusivity should be at the forefront of engineering concepts and railway design. However, despite the ever-modernising railway infrastructure, innovative, inclusive design is prioritised for Network rail's 'taxpayers, customers and passengers' (Network rail, 2019).

With the construction of some railway depots dating back to Victorian times, there is little consideration for inclusive design in depots, leaving many workers to feel excluded and unemployable as the already 'narrow specification of an '*ideal candidate'* immediately discounts many disabled people' (Libby, 2019).

However, what happens when a non-disabled member of staff injures themselves?

Many depot workers lose their lives or become severely injured whilst at work (Cant, 2012; Laskow, 2018), and, with little-to-no accessibility or inclusive design elements within railway depots, they are unable to return safely to their workplace (Glasswall, 2007; RAIB, 2007b, 2007a, 2007c, 2020, 2021; Office of Rail and Road, 2018; Spence, 2019; Stewart and RSSB, 2019; Horgan, 2020; Pitt, 2020; Bradshaw, 2021; Iosh, 2021). Because of this, 'safety inspectors are demanding Network Rail implement "real change"' (Topham, 2020) to minimise the fallout of inadequate conditions that railway workers face, reducing the unemployment of disabled workers and increasing the working potential for those who attain a life-changing injury whilst at work. Whilst the ideology of inclusivity is a broad concept, there is a spectrum of ailments to design for, whether the situation is permanent, temporary or situational, as denoted by the persona spectrum in Figure 1.

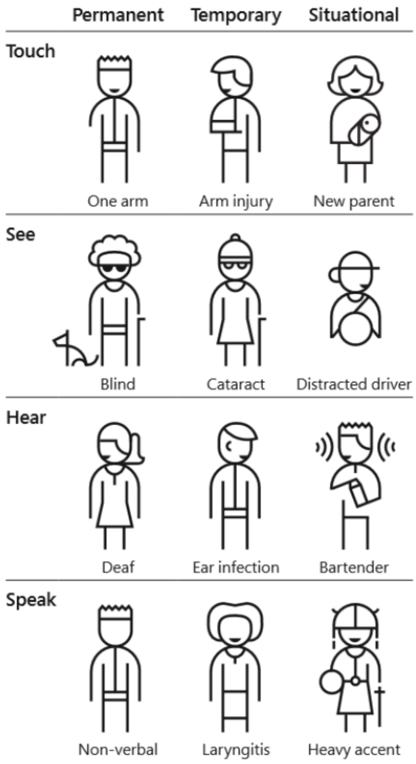


Figure 1: The persona spectrum – identifying permanent vs situational disabilities (Human Machine Interface Expert, 2017)

Inclusive design is perhaps the most up-and-coming trait of design engineering in rail. Now, in an ever-inclusive and all-involving world (Network rail, 2015), exclusivity should be a thing of the past. Today sees a new league of railway workers, including women, non-UK-natives, the disabled and the elderly, coming together in unison to form a vast and multifaceted workforce (Gillham, Thomas and Jake, 2021).

The risk of injury for those who work in railway depots is 'irrefutable but should not be inevitable' (Cortes, 2021). The implementation of faster trains, higher voltage electricity and ‘powerful machinery combines to make modern depots potentially deadly places to work.’ (RailEngineer, 2021). This is further compounded by the ‘growing number of vehicles on the network, leading to mounting pressure on operators and a desire to achieve ‘pitstop style’ servicing’ (RailEngineer, 2021). Although staff are hired and trained to a competency level deemed adequate by the ORR (office of rail and road), often they must complete a fitness regime which encompasses mental fitness, medical fitness, and physical fitness. These are defined as:

- Physical fitness: an individual ‘possesses the physical attributes of strength, agility etc., enabling the activity to be performed competently and safely’ (Cleeton, 2011).
- Mental fitness: ‘implies that no existing mental conditions may adversely affect concentration, decision making or behaviour, and so compromise competence and safety’ (Cleeton, 2011).
- Medical fitness: ‘covers any medical condition that may adversely affect competence and safety at present or in the future’ (Cleeton, 2011).

As companies are legally not allowed to discriminate against an individual in the workplace, they ‘should consider whether activities can be adapted to enable those with physical or medical limitations to work’ (Cleeton, 2011). Not doing so creates ‘unfair barriers for employment’ (Cleeton, 2011), reducing the intake of new employees and disallowing injured employees to continue their employment.

Despite rail staff retaining injuries from the workplace, many employees find themselves unable to return to work after their accident due to the inability to operate machinery or mental health concerns. Staff have gone to press stating that ‘Network Rail [have] not adequately addressed the protection of track workers’ (Horgan, 2020), adding, ‘we [are] being asked to work in incredibly unsafe conditions… my accident could have been prevented’ (Castle, 2021). Furthermore, although lessons learnt from accidents are recorded to be prevented, few design iterations are made in response to incidents, meaning that those who have suffered a life-changing injury cannot return to the same pre-accident workplace.

The Approach

After preliminary research and discussions with railway workers about issues they face with accessibility in the workplace and returning to work after an injury, a questionnaire was formed to determine the most significant factor affecting their ability to do their job. Train operating companies, management and depot workers were interviewed on what they felt were the essential aspects of depot design to provide an inclusive environment. This created a list of attributes about which to ask employees.

The work was sent to railway depot workers in the UK to capture representative information via an online survey. In the survey, workers were asked to rank the importance of different inclusive design features concerning how significantly not having them would impact their job. Data was compiled, and participants were re-interviewed to discuss any concerns regarding their findings in the depot.

Once the data was collected, it was ranked and given a relative importance statistic so that analysis of depots could be undertaken as a ‘tick box’ style exercise for depot designers.

Findings

This research found that staff value toilets, bright light working conditions, clear walkways and washing facilities over better coffee machines, mental health assistance and paid breaks. Despite the preliminary research branching into accessible design solutions for those injured at work who cannot return, findings showed a significant issue with current infrastructure and non-injured workers.

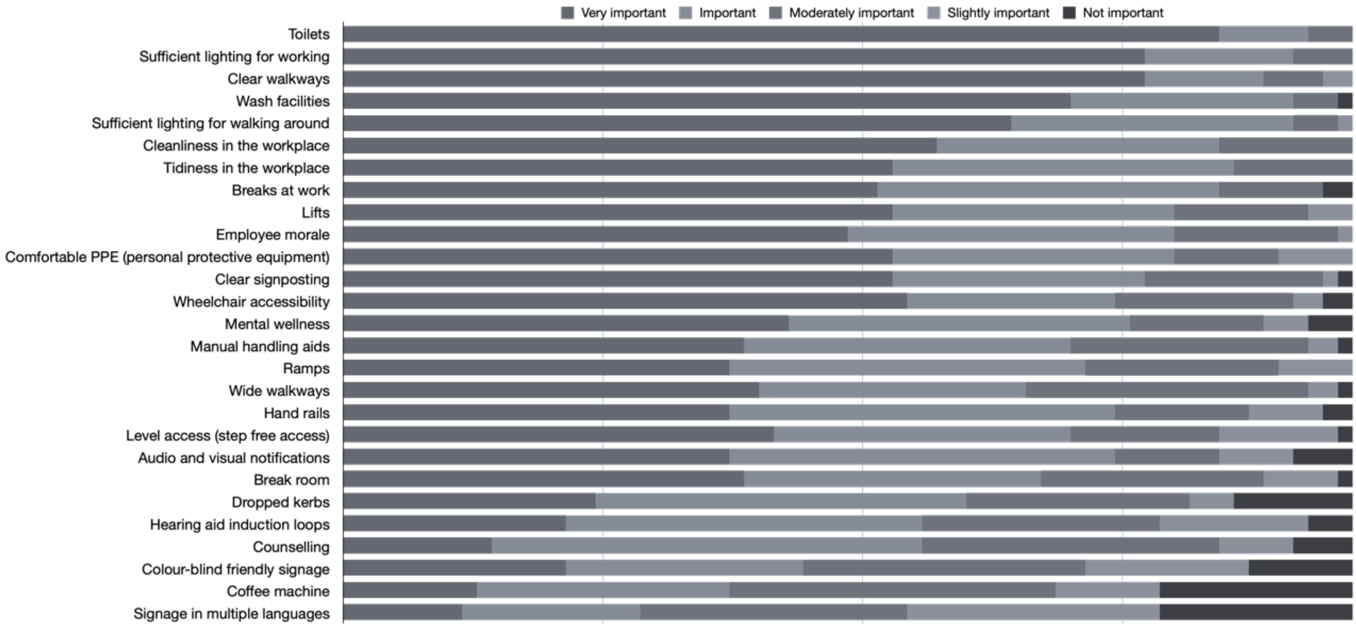


Figure 2: Graphical representation of votes cast against inclusive importance factors.

Workers value toilet facilities in the workplace far above any other characteristic in depot design. By contrast, the least essential attributes in creating an inclusive environment, according to depot workers, were multiple language signing, coffee machines and counselling.

Interestingly, feedback from the industry places coffee machines and multiple-language signposting as equally unimportant. Despite the employment of ethnic minority workers, multiple language signposting is not valued highly and was deemed the lowest in terms of importance. This could be, in part, due to a smaller percentage of ethnic minority staff members or due to the inability to reach out further due to the pandemic limiting social interaction.

Data extraction

Data captured from the questionnaire enabled qualitative analysis of depot inclusivity but did not provide the ability to analyse depots quantitatively, thus making the study susceptible to bias by personal interpretation; extracting the data from the questionnaire and mapping the feedback to a quantitative figure provided the basis for analysing depots quantitively and minimalising the risk of error or miss-elucidation.

A matrix was drawn to deduce quantitative data from the results, giving the results from the questionnaire a score; 1 to 5, in conjunction with the importance spectrum in the questionnaire. The resulting matrix was designed whereby the qualitative result *Very important* was given the integer 5, and the lowest end of the spectrum was given the integer 1 for *unimportant*.

The matrix was then fitted to the data from the questionnaire, which provided a score, out of a possible 340 (68 participants providing a score out of 5 points per item), for the importance rating of inclusive aspects of depot design.

To quantify any future depot designs in respect to their inclusivity, data from the questionnaire was quantitated, and then, a weighting factor was produced using the ratio of importance against total possible importance.

'Statistical weight is an amount given to increase or decrease the importance of an item' (Glen, 2019). In this instance, the importance of an inclusive aspect could be calculated by dividing the total possible score of importance by the actual score of importance.

Table 1: Creating a weighting factor using data from the questionnaire

Results from inclusive design questionnaire (decending order)

Question	Total score for inclusivity	Amount of votes					Weighting factor	Weighting factor * 10	Score out of 10 for weighting factor
		Very important	Important	Moderately important	Slightly important	Not important	340/total score		
Toilets	328	59	6	3	0	0	0.96	9.65	9.6
Sufficient lighting for working	322	54	10	4	0	0	0.95	9.47	9.5
Clear walkways	318	54	8	4	2	0	0.94	9.35	9.4
Wash facilities	315	49	15	3	0	1	0.93	9.26	9.3
Sufficient lighting for walking around	312	45	19	3	1	0	0.92	9.18	9.2
Cleanliness in the workplace	303	40	19	9	0	0	0.89	8.91	8.9
Tidiness in the workplace	301	37	23	8	0	0	0.89	8.85	8.9
Breaks at work	295	36	23	7	0	2	0.87	8.68	8.7
Lifts	294	37	19	9	3	0	0.86	8.65	8.6
Employee morale	293	34	22	11	1	0	0.86	8.62	8.6
Comfortable PPE (personal protective equipment)	292	37	19	7	5	0	0.86	8.59	8.6
Clear signposting	292	37	17	12	1	1	0.86	8.59	8.6
Wheelchair accessibility	288	38	14	12	2	2	0.85	8.47	8.5
Mental wellness	278	30	23	9	3	3	0.82	8.18	8.2
Manual handling aids	276	27	22	16	2	1	0.81	8.12	8.1
Ramps	275	26	24	13	5	0	0.81	8.09	8.1
Wide walkways	274	28	18	19	2	1	0.81	8.06	8.1
Hand rails	273	26	26	9	5	2	0.80	8.03	8.0
Level access (step free access)	272	29	20	10	8	1	0.80	8.00	8.0
Audio and visual notifications	271	26	26	7	5	4	0.80	7.97	8.0
Break room	271	27	20	15	5	1	0.80	7.97	8.0
Dropped kerbs	244	17	25	15	3	8	0.72	7.18	7.2
Hearing aid induction loops	242	15	24	16	10	3	0.71	7.12	7.1
Counselling	240	10	29	20	5	4	0.71	7.06	7.1
Colour-blind friendly signage	225	15	16	19	11	7	0.66	6.62	6.6
Coffee machine	206	9	17	22	7	13	0.61	6.06	6.1
Signage in multiple languages	189	8	12	18	17	13	0.56	5.56	5.6

The weighting factor was calculated as a whole integer instead of a percentage, as research suggests that percentages can be misleading because it is 'hard to know if the percentage was calculated using the original numbers or the total resulting from the change' (Krause, 2017). Secondly, they were calculated as a whole number for ease of addition when adding together the inclusive aspects of a depot during the analytic phase of the research.

Results – how we can use this in the future

Feedback from the questionnaire enabled a qualitative study to be evaluated quantitively, allowing a hierarchical study of the most important aspect of inclusive design against inclusive design elements that perhaps were not as important. Results showed that toilets were deemed the most crucial inclusive design product, with coffee machines being one of the least important. With this, the results were weighted using a statistical weighting factor which determined the worth of every piece of inclusive design in the questionnaire.

A handful of depots were picked for demonstrative purposes, and it was determined that the least inclusive depot was Aylesbury (at 36%), with the most inclusive being Hitachi's depot in three bridges (at 66%). Hopefully, this method could be applied to depots throughout the UK to create a tolerable and intolerable region for inclusion, justifying expenditure for bettering rail depots with cost-benefit analyses and user-centred design iterations.

Table 2: An example of how a depot can be quantitively assessed to distinguish its inclusivity (DC, 2012; Marshall, 2017; Thorkildsen, 2017b, 2017a; Chiltern Railways, 2019)

Name	Weighting factor /10	Aylesbury Maintenance depot	Northumberland Park Depot	Three bridges depot	Stoke Gifford depot	Slade green depot
Toilets	9.6	✓	✓	✓	✓	✓
Sufficient lighting for working	9.5	☐	✓	✓	✓	✓
Clear walkways	9.4	✓	✓	✓	✓	✓
Wash facilities	9.3	☐	☐	☐	✓	☐
Sufficient lighting for walking around	9.2	✓	✓	✓	✓	✓
Cleanliness in the workplace	8.9	✓	✓	✓	✓	✓
Tidiness in the workplace	8.9	✓	✓	✓	✓	☐
Breaks at work	8.7	✓	✓	✓	✓	✓
Lifts	8.6	☐	✓	☐	✓	☐
Employee morale	8.6	☐	☐	☐	☐	☐
Comfortable PPE (personal protective	8.6	✓	✓	✓	✓	✓
Clear signposting	8.6	☐	☐	✓	☐	☐
Wheelchair accessibility	8.5	☐	☐	✓	☐	☐
Mental wellness	8.2	☐	☐	☐	☐	☐
Manual handling aids	8.1	☐	☐	☐	✓	☐
Ramps	8.1	☐	☐	✓	✓	✓
Wide walkways	8.1	✓	✓	✓	✓	✓
Hand rails	8.0	☐	✓	✓	✓	✓
Level access (step free access)	8.0	✓	☐	✓	✓	☐
Audio and visual notifications	8.0	☐	☐	✓	☐	☐
Break room	8.0	☐	✓	✓	✓	☐
Dropped kerbs	7.2	☐	☐	✓	☐	✓
Hearing aid induction loops	7.1	☐	☐	☐	☐	☐
Counselling	7.1	☐	☐	☐	☐	☐
Colour-blind friendly signage	6.6	☐	☐	☐	☐	☐
Coffee machine	6.1	☐	☐	☐	☐	☐
Signage in multiple languages	5.6	☐	☐	☐	☐	☐
Total score	220	79.26	105.38	145.03	138.85	95.18
		35.99%	47.84%	65.84%	63.04%	43.21%

However, these are only illustrative figures; with the coronavirus outbreak, unauthorised personnel could not visit railway depots, so the results come from online research and videography of depots and should only be considered trial numbers.

This report found that staff value toilets, bright light levels, clear walkways and washing facilities over coffee and paid breaks. To achieve full working potential, staff want to change in the infrastructure, not mental health change. Adding lifts, larger car parks, and tidier workplaces are more desired than counselling or well-being to employees, which is interesting considering the multitude of well-being discussions within the industry today.

Toilets were the most desirable asset that track workers wanted as often these facilities are not provided when trackside. Staff working trackside often do so for almost 12 hours, limiting the amount they drink, so they do not need to go to the toilet. Additionally, it is often assumed that only men work trackside; therefore, toilets are not needed due to their ability to go elsewhere. The consensus from female entries in the questionnaire is that there is very little consideration for a woman's menstrual health. Often, toilets are locked, used as storage facilities, and do not provide any menstrual products or sanitary waste bins where they are needed.

Discussion

By far, the most prevalent finding for the inclusive design in railway depots was not that there was a significant issue with accessibility requirements, nor that staff felt they deserved better equipment, but simply the lack of toilet facilities in depots and trackside. Despite case study research instigating preliminary studies into accessibility and inclusivity of depots through conventional design choices, conversations with those working in railway depots inferred a less typical story. Although there may be inaccessible working environments for injured people, there are also inclusive elements missing from depot infrastructure that enable comfortable working practices.

Despite the push for more female engineers, it was discovered that women often have no toilets when trackside, no sanitary waste bins for disposal of menstrual products, and must wear clunky men's shoes and men's high visibility jackets when working. Discussions with female depot workers unearthed stories of embarrassing accidents when working, which could otherwise have been avoided by implementing adequate sanitary waste bins and toilets.

Although there is a much smaller percentage of female engineers and track workers than males, inconsistencies with sanitary waste and appropriate PPE discourage many women from working in the environment. A vicious circle, perhaps implementing or sourcing portable toilets, could solve many of the workers' concerns.

Furthermore, as societal views around cleanliness and hygiene from the coronavirus pandemic alter, railway depots will also change. Data shown in this study is not necessarily reflective of a post-pandemic world but rather a snapshot of mid-pandemic lifestyle preferences.

Perhaps most interesting from the findings was the lack of importance of 'typical' inclusive design characteristics such as multiple language signage, step-free access and audio and visual notifications. This could be partly due to the small percentage of workers who may require these additional ailments and have their vote lessened to some degree. However, for future work, it would be interesting to capture the requirements again using mobility and sensory ailment simulators to see how these considerations may change when mimicking returning from having an accident at work.

Study limitations

The work conducted for this paper was developed and carried out during the coronavirus pandemic. Abiding by the government's stay-at-home policy, all research for this paper was done remotely. In an ideal world, station depots would have been researched in person, and workshops and events held to appropriately capture a more extensive breadth of knowledge about railway depots and their workers. Furthermore, implementations could have been carried out, and a more representative study could have been performed. Due to this, the findings from this paper are merely representative and provide a possible methodology for capturing depot inclusion which could be implemented and adapted to the workers' environment changes.

Additionally, though the project was targeted at being inclusive through the study, the limitations in being able to sit face-to-face with people meant that there are likely proportions of the population that would have been unable to complete the survey and answer questions. The inability to run workshops meant that open discussions were also limited, making it harder to capture data from casual conversations with the workers.

Conclusions

Railway depots are complicated systems with various tasks and deliverables that must be undertaken to achieve goals within designated timescales. Due to the fast-paced nature of the work, it is not uncommon for staff to acquire injuries which, unlike in other practises, leave them unable to continue their job. This, combined with seemingly unclean and inaccessible working environments, leaves staff unhappy and wanting change.

With the everchanging demographic of society and the modernisation of medicine, design engineers are tasked more heavily than ever to construct pieces that suit a wide-ranging audience. Nowadays, illness and medical conditions are rarer and rarer, so anthropometric data or traditional ergonomics cannot quantify them.

Whilst inclusive design is somewhat of a novel concept in railway depots, quantifying how inclusive infrastructure is would massively reduce bias in depot analysis. This would allow for much more proficient and high-end design iterations, making design engineers focus on user-centred design studies rather than anthropometric data and previous design iterations. Quantifying inclusivity also would ensure that depots would be in keeping with the current needs of the public; with a database that could be ever-growing, the weighting factors for importance in inclusive design elements would reflect the current situation in the public eye. It would also see a prolific change in how railway infrastructure is designed, putting inclusivity at the forefront of any engineering concept rather than an iterative design added at a later stage.

References

Bradshaw, K. (2021) *'Every day is a struggle': Rail worker who lost arm in accident calls for freight firm to apologise*. Available at: https://www.itv.com/news/meridian/2021-05-25/every-day-is-a-struggle-rail-worker-who-lost-arm-in-accident-calls-for-freight-firm-to-apologise (Accessed: 29 June 2021).

Cant, D. (2012) *One of the most dangerous jobs in the working world*. Available at: https://www.veritas-consulting.co.uk/blog/one-of-the-most-dangerous-jobs-in-the-developed-world-working-on-the-rail-tracks/ (Accessed: 5 July 2021).

Castle, L. (2021) *Shunter driver Terry Currie from Maidstone who lost arm in horrific crash in Dollands Moor Freight Yard in Folkestone demands employer DB Cargo admits responsibility*. Available at: https://www.kentonline.co.uk/maidstone/news/man-hit-by-freight-train-at-work-demands-employer-takes-resp-247901/ (Accessed: 29 June 2021).

Chiltern Railways (2019) *Chiltern Railways Short Form Update - A look inside Aylesbury Maintenance Depot*. UK: YouTube.

Cleeton, G.U. (2011) 'Fitness for work.', *Making work human.*, (December), pp. 101–127. Available at: https://doi.org/10.1037/13246-005.

Cortes, M. (2021) *Rail worker deaths should not be inevitable: It's time to put the brakes on slipping rail safety*. Available at: https://www.newcivilengineer.com/latest/rail-worker-deaths-should-not-be-inevitable-its-time-to-put-the-brakes-on-slipping-rail-safety-15-02-2021/ (Accessed: 8 July 2021).

DC, J. (2012) *(HD) A Look Inside Northumberland Park Depot On The Victoria Line*. UK: YouTube.

Gillham, J., Thomas, T. and Jake, F. (2021) 'Ethnicity Pay Gap Report 2021', (March 2019). Available at: https://www.strategyand.pwc.com/uk/en/reports/ethnicity-pay-gap-report.pdf.

Glasswall (2007) *Accident at leatherhead*. Available at: https://www.gov.uk.glasswall-icap.com/raib-reports/accident-at-leatherhead (Accessed: 2 July 2021).

Glen, S. (2019) *Statistics how to*, *Probability and Statistics Topic Index*. Available at: https://www.statisticshowto.com/weighting-factor/#:~:text=To calculate how much weight,) %3D 49%2F59 %3D . (Accessed: 5 August 2021).

Horgan, R. (2020) *Track worker safety neglected 'for many years' ahead of Margam tragedy*. Available at: https://www.newcivilengineer.com/latest/track-worker-safety-neglected-for-many-years-ahead-of-margam-tragedy-12-11-2020/ (Accessed: 30 June 2021).

Human Machine Interface Expert (2017) *Inclusive Design at Microsoft*.

Inclusive design hub (2021) *Inclusive design*. Available at: https://inclusivedesign.scot/what-is-inclusive-design/ (Accessed: 7 May 2021).

Iosh (2021) *DB CARGO FINED £200,000 OVER SHUNTER'S ARM AMPUTATION*. Available at: https://www.ioshmagazine.com/2021/03/23/db-cargo-fined-ps200000-over-shunters-arm-amputation (Accessed: 30 June 2021).

Johnson, T. (2023) 'Rail worker injuries increase in last year despite safety reforms', *Paper Knowledge . Toward a Media History of Documents*, pp. 12–26. Available at: https://www.newcivilengineer.com/latest/rail-worker-injuries-increase-in-last-year-despite-safety-reforms-09-01-2023/.

Krause, H. (2017) *WHY PERCENT CHANGE IS ACTUALLY MISLEADING (MOST OF THE TIME)*. Available at: https://idatassist.com/why-percent-change-is-actually-misleading-most-of-the-time/ (Accessed: 13 August 2021).

Laskow, S. (2018) *Railyards Were Once So Dangerous They Needed Their Own Railway Surgeons*. Available at: https://www.atlasobscura.com/articles/what-did-railway-surgeons-do (Accessed: 5 July 2021).

Libby (2019) *From expert to unemployable*. Available at: https://www.leonardcheshire.org/our-impact/stories/expert-unemployable (Accessed: 10 April 2021).

Marshall, G. (2017) *Inside GWR Hitachi Depot*. UK: YouTube.

Network rail (2015) 'Spaces and Places for Everyone'.

Network rail (2019) *Spaces and places for everyone*. Available at: https://cdn.networkrail.co.uk/wp-content/uploads/2019/04/Spaces-and-Places-for-Everyone.-Our-Inclusive-Design-Strategy.pdf (Accessed: 7 May 2021).

Office of Rail and Road (2018) *Prosecution*.

Pitt, E. (2020) *'There's no sense to what happened': Relative of Margam train victim speaks of 'heartbreak' a year after deaths*.

RAIB (2007a) *Accident at Leatherhead*.

RAIB (2007b) *Fatal collision between a super Voyager train and a car on the line at copmanthorpe*.

RAIB (2007c) 'Rail Accident Report Collision between a train and a road vehicle ', (September).
RAIB (2020) 'Track workers struck by a train at Margam Neath Port Talbot', (July).
RAIB (2021) *Collision between RRVs at Ramsden Bellhouse*. Available at: https://www.gov.uk/government/news/collision-between-rrvsat-ramsden-bellhouse (Accessed: 28 June 2021).
RailEngineer (2021) *Safety by design*. Available at: https://www.railengineer.co.uk/safety-by-design/ (Accessed: 28 June 2021).
Spence, A. (2019) *Electrical fire - staff injury*. Available at: https://safety.networkrail.co.uk/wp-content/uploads/2019/01/Safety-Bulletin-NRB-19-01-Godinton-Substation-Staff-Injury.pdf (Accessed: 5 July 2021).
Stewart, O. and RSSB (2019) 'Collision between a train and utility vehicle at Dollands Moor freight yard, Kent', 12(February).
Thorkildsen, A. (2017a) *Inside Slade Green Railway Depot*. UK: YouTube.
Thorkildsen, A. (2017b) *Inside Three Bridges Railway Depot*. UK: YouTube.
Topham, G. (2020) *Calls for 'real change' at Network Rail after track workers' deaths*. Available at: https://www.theguardian.com/uk-news/2020/nov/12/calls-real-change-network-rail-track-workers-deaths-wales (Accessed: 5 July 2021).
World Health Organisation (2020a) *Disability*. Available at: https://www.who.int/health-topics/disability#tab=tab_1 (Accessed: 7 May 2021).
World Health Organisation (2020b) *Disability and Health*. Available at: https://www.who.int/news-room/fact-sheets/detail/disability-and-health (Accessed: 7 May 2021).

Performance and workload using an audible intelligent assistant during pilot training

David A. Hudson & Michael A. Bromfield

University of Birmingham

SUMMARY

Boeing forecast that globally an additional 602,000 pilots will be needed by 2041 to meet year on year growth of 3.6% in passenger traffic. All pilots need to be trained in accordance with accepted regulatory standards. New technology is continuously being developed to enhance training and reduce training time. Research into the effectiveness of training technologies and how these impact pilot performance and workload is key to future growth. One such technology under review at the University of Birmingham is the 'audible intelligent assistant'. This artificial voice that provides real-time feedback to the pilot during flight training. The system provides warnings, cautions and instruction to the human pilot to enhance pilot training. Preliminary results of the research showed an improvement in pilot performance against a specified set of target parameters of airspeed and altitude and a corresponding decrease in workload for 80% of pilots when using the audible intelligent assistant. Without AIA, pilot performance improved by 35% due to learning effects alone, with AIA pilot performance showed an improvement of 65%.

KEYWORDS

Pilot Training, Audible Intelligent Assistant, Performance, Workload

The Experiment

A range of experiments were conducted to determine the effects on Pilot Performance and Workload using the audible intelligent assistant (AIA). Ten participants with real-time flying experience between 5 and 50 hours and were randomly selected from University of Birmingham students. All participant's data were deidentified and aggregated and followed the University's ethical procedures. The mean number of hours flying being 9 and mean age of the participants was 20 across the cohort of 10 participants. The participants were split allocated to two groups. Group 1 - the control group - who performed flight exercises without the audible assistant, and Group 2 - the experimental group – who performed exercises with the AIA. During the experiments, the AIA (based on gaming technology) was combined with a basic fixed-base flight simulator to provide audible instructions, cautions and warnings to the pilot based on their performance of the flying task in the circuit. The amount of verbal communications was standardised however more erratic flying would result in more instruction, cautions and warnings. Indicated performance parameters were used to determine the pilot's variation from the target and to analyse the pilot's performance within a set completion standard set out by the FAA in accordance with the suitable aircraft parameters for the Cessna C172. A NASA-TLX survey was used post-flight to assess Pilot Workload for the given flying tasks.

Analysis

The preliminary results of this study indicated that pilots in the experimental group demonstrated greater performance increases while using the AIA, with 26 out of 40 parameters (65%) across the 5 pilots. Four out of five of the pilots in the experimental group showed a reduction in workload compared to the control group. With respect to the measured performance parameters of altitude, airspeed and heading during the final approach, the experimental group showed significant improvements between the first and second circuits. However, both experimental and control groups struggled to show a clear improvement in performance in the crosswind, downwind and base leg of the flight. There were marginal improvements in performance between the two groups of pilots for the take-off (heading and bank angles) with the use of the AIA. The control group (not using the aid) showed a decrease in performance for 24 out of 40 parameters (60%) across the all 5 pilots and an increase in workload for 4 out of 5 of the group. The parameters recorded being airspeed, altitude, and average bank angle as well as 5 different headings for each turn of the circuit. The preliminary results suggest that use of such an audible tool can simultaneously improve performance and decrease workload leading to safer and more cost effective pilot training. The perceived increase in workload and recorded decrease in performance for the control group between circuits may have been due to self-reflection and self-induced pressure to demonstrate improvement, without feedback. The AIA provided clear instruction prior to flight about the aircraft's capabilities for a turn but also, warnings during flight. For example, the yaw effects that an single engine propeller driven aircraft encounters due to the application of full power for the take-off run may be countered by prompts from the intelligent audible assistant. However, this instruction usually came too late during the take-off run and yawing was evident hence no clear improvement. The audible assistant was constantly providing instructions based on airspeed and altitude benefitting the pilot's performance and showing a clear improvement and stability of the engine power. The reason for both groups showing significant improvement on the final approach heading was due to pilots relying upon their cognitive skills and using the information provided in the briefing to point out a visual landmark that could be used as a turning point (to compensate for the limited horizontal field of view in the simulator). A lack of improvement in heading between the two groups is partly due to the lack of communication from the assistant, potentially also due to the coarse divisions of the compass heading system used (5 deg.), thus the required headings lie between these values and made reading subjective. Furthermore, the tone at which this audible instruction is provided can have a significant effect on performance, with pilots benefitting more positively in terms of performance from negative or neutral tones. The audible software gave no positive reassurance or praise for a pilot’s accomplishments in any scenarios and tended only to highlight mistakes or the need for improvements. Also considered are how the software displays the four key aspects of feedback (Molesworth et al., 2006; Molesworth et al., 2011) such as when, how often and what feedback is provided as well as how much consideration is needed to execute the feedback.

Impact & implications

While the results of the research highlight the potential benefits for training, all the pilots had different simulator experiences which may have affected results slightly. Despite reasonable fidelity of the simulator, 100% replication of real flying experience was not possible. The results may have been influenced by the pilot’s individual learning preferences either visual or auditory learners (Chui et al, 2020). During basic student pilot training typically, a flight instructor would provide audible instructions and crucial training points however the information provided is sometimes inconsistent within the flight school environment. Using an audible assistant in training could improve standardisation and suggests that pilot performance can be improved, and pilot workload simultaneously reduced for the limited parameter set analysed. These changes could have a beneficial impact on student pilot training leading to greater efficiency and safety. When not using AIA, performance decreased, and workload increased. Further experiments are planned with increased participants, but initial signs are encouraging and use of an AIA may be a first step towards improving the quality and consistency of flight training whilst also reducing cost.

References

Boeing (2022), Pilot and Technician Outlook, 2022-2041, https://www.boeing.com/commercial/market/pilot- technician-outlook/

Chui, T, Molesworth, B., & Bromfield, M., (2020), "Feedback and Student Learning: Matching Learning and Teaching Style to Improving Student Pilot Performance", International Journal of Aerospace Psychology.
https://doi.org/10.1080/24721840.2020.1847650

Molesworth, B. R. C., Bennet, L., & Kehoe, E. J. (2011). Promoting learning, memory, and transfer in a time- constrained high hazard environment. *Accident Analysis and Prevention, 43, pp* 932- 938.

Molesworth, B. R. C., Wiggins, M., & O'Hare, D. (2006). Improving pilots' risk assessment skills in low-flying operations: The role of feedback and experience. *Accident Analysis and Prevention, 38, pp* 954 – 960.

Digitalisation of HFE in Medical Product Development: Challenges and Opportunities

Diego Cortez & Erin Davis

Emergo by UL Solutions

SUMMARY

This paper presents current challenges of applying human factors engineering (HFE) throughout medical device development, and the opportunity digitalisation creates for innovation in the field. The paper focuses on describing current HFE challenges and examples of how digital tools and software applications can contribute to the work of HF specialists in the medical industry, noting that there is large, unmet need for more HF expertise. Finally, it briefly presents a case study of a software released by Emergo by UL Solutions which aims to address some of these challenges.

KEYWORDS

Human factors, digitalisation, healthcare

Content

Presently, there are approximately 600,000 medical devices available on the UK market (Department of Health & Social Care, 2021) with approximately 806 medical manufacturers (Bold Data, 2022). However, there is nowhere near that many HF specialists, even assuming one person per company was sufficient. The Chartered Institute of Ergonomics & Human Factors (CIEHF) lists 513 registered members and fellows, many of which work in multiple industries beyond just the medical industry. Considering the growing demand of rigorous HFE driven by regulators of the biggest medical device markets such as the US, the UK, the EU, and China, there are not enough HFE/medical specialists to meet the industry's needs. The insufficient supply of HFE specialists and the increasing demand for more HFE expertise encourages the development of digital solutions.

Digitalisation could address some of the challenges that medical device manufacturers face when integrating HFE into the development of their products. First, digitalisation can contribute to teaching specialists and non-specialists about basic and, in some cases, more advanced and nuanced aspects of applying HFE to medical technology. Second, digitalisation can provide practitioners with a framework for performing HFE work in a complete and effective manner that will address today's regulatory and commercial imperatives. Third, digitalisation can provide practitioners with productivity tools as well as tools that help them produce innovative and high-quality results, including essential analyses, user interface designs, and evaluations.

Oversimplifying the HFE process in medical devices into three steps (i.e., concept, design, evaluation) for didactic purposes, Table 1 illustrates current HFE challenges and digitalisation opportunities for each step.

Table 1: HFE Challenges and Digitalisation opportunities

HFE Step	Today's Challenges	Future digital solutions
Concept	Uncertainty regarding regulatory requirements driving certain types of HFE projects	Digital checklists, tools, and workflows that guide users through key decisions and recommend HFE activities
Concept	Finding relevant results in a known problems analysis; reviewing unstructured and inconsistent data	Machine learning and/or artificial intelligence enabled data analysis
Design	Lack of special-purpose and convenience tools to assess whether design meets HFE design principles (e.g., text legibility)	Digital calculators that translate HFE principles into actionable design recommendations
Evaluation	Drawing upon supporting HFE principles during root cause analysis of use errors and appropriate root cause analysis descriptions	Digital libraries that present easy-to-leverage HFE principles that inform root causes
Evaluation	Writing clear and appropriate root cause analysis descriptions	Digital libraries of common root cause analysis descriptions
Evaluation	Analysing large, unwieldy data sets arising from large-sample usability tests	Automated data transfer from datasheet to HF validation report tables
All	Challenges performing nuanced, intensive HFE tasks (e.g., use-related risk analysis, residual risk analysis)	Digital tools that guide teams through key analyses and enable them to evaluate the quality of HFE deliverables
All	Due to lack of awareness of HFE, HFE experts must first convince management on the value of HFE before sufficient resources are allocated to get the work done	Help all stakeholders understand HFE by creating organisation-wide awareness of why and how to perform HFE using e-learning training packages.

Recognising these challenges, Emergo by UL Solutions embarked in an intensive two-year research and development effort involving dozens of medical product manufacturers and HFE subject matter experts from the US, the UK, The Netherlands, and Japan. Emergo's HFE platform the Optimal Product Usability Suite (OPUS™), matches a trend toward the use of digital tools to facilitate the work of HFE specialists. OPUS addresses common HFE challenges by a) being targeted towards users working in the regulated medical industry to meet regulators' guidance and industry standards (e.g., IEC 62366-1:2015); b) providing self-paced, online training regarding the specifics of applying HFE for medical devices subject to regulatory approval; c) providing helpful tools to support common research and design analyses; and d) ensuring that practitioners have "checked all the boxes" when developing a medical device subject to regulatory approval.

For the most part, digital solutions will not replace human experts. Rather, the solutions will complement and augment the work of experts. As such, human factors experts should explore opportunities to create digital solutions that will enable them to work efficiently and focus their efforts on the most demanding HFE tasks that only a human can perform.

References

Department of Health & Social Care. (2021). Policy Paper Factsheet: medical devices overview. https://www.gov.uk/government/publications/medicines-and-medical-devices-bill-overarching-documents/factsheet-medical-devices-overview

Bold Data. (2022). Medical Device Companies UK. Retrieved from: https://bolddata.nl/en/companies/uk/medical-device-companies-uk/

Review of UAV Loss of Control In-Flight: Accidents and Incidents

Rahma El Safany[1], Michael A.Bromfield[2]

[1] Master's in Aerospace Engineering student, University of Birmingham, [2] Associate Professor in Aerospace, University of Birmingham. CIEHF Member & Corresponding Author.

SUMMARY

Sixty Uncrewed Aerial Vehicle accident reports were analysed to identify possible causal and contributory factors leading to loss of control in flight and recovery actions where applicable. Manufacturing and design errors were dominant in 22 causal factors (34% of events) and 18 contributory factors (22% of events) (e.g. ingestion of precipitation). Recovery was not attempted in 35 (55%) events. The relationship between age, total hours experience, hours experience on type, recovery attempts and number of accidents increase with operator age or lack of experience was also analysed. As total experience increases the number of accidents and attempted recovery increases. All this information is presented in a framework adapted from the Accident Route Matrix to recognise loss of control in flight in future accidents and improve recovery response.

KEYWORDS

UAV, LOC-I, accident, Causal Factors, Contributory Factors

Introduction

Loss of control in flight (LOC-I) is the most frequent and significant cause of accidents for commercial and general aviation (IATA, 2019). LOC-I has been recently re-defined for both commercial (Bromfield & Landry, 2019) and general aviation (Smith & Bromfield, 2022) but no definition exists for UAVs. Despite accidents involving uncrewed aerial vehicles (UAVs) being 30 times higher than crewed aerial vehicles (McCarley et al), the causes of these accidents have not been previously investigated. The aim of this research is to identify the main causal and contributory factors leading to LOC-I for UAVs, considering human factors, automation levels and recovery methods to attempt to regain control. Sixty UK Air Accident Investigation Board civil aircraft accident and serious incident reports for UAVs within the UK, were analysed to help define LOC-I for UAVs and provide an illustrative framework for accident analysis and consistency in reporting.

Preliminary Analysis

Pre-conditions are operating conditions that do not alter before or during flight but may have an impact on commander's response in case of an upset. These include commander's age and experience and the type of operation. Each accident was thoroughly analysed to identify and categorise the main causal and contributory factors. The causal factors were identified as actions, omissions, events, conditions, or a combination thereof, that led to an accident or incident (McCarley et al., 2004). All other events after the primary causal event were considered as contributory factors. The accidents were categorised through a normalisation process based on their similarity and these categories were analysed using statistics. The main causal factors (34%) and contributory factors (22%) leading to LOC-I for UAVs were manufacturing failures. Issues related to human factors in UAVs were based on the interaction of the operator with the aircraft since they are not co-located (McCarley et al.,2004). To analyse the effect of human error on UAV operations, the main factors considered with respect to the number of accidents were: recovery methods attempted after LOC-I, the level of UAV automation, the operator's interface and reliance on autonomous features, the operator's age, total hours of experience/experience on type. Only three of the total number of accidents where recovery was attempted (45%) were successful. In 55% of accidents recovery was not attempted and this was due to insufficient time, the pilot losing sight of the UAV or not recognising LOC-I due to warning messages not being displayed. This infers pilot's inherent trust in automation and high reliance on 'fail-safe' functions and warning messages to recognise LOC-I. Previous literature suggests that most human errors are caused by design inconsistencies of the ground control station (GCS) and failure in autonomous devices to predict or respond to all scenarios (Nilsson, 2011). The results of the accident analyses suggest that lower age groups (20-29 years) may rely more on automation during recovery. The largest reliance on 'return to home' and 'kill switch' functions was found for operators with less than 100 hours of experience. This may also be linked to lack of knowledge on how to use the equipment provided and/or low situation awareness. The results suggest that as experience on type increases, situation awareness and readiness of the pilot may also increase, helping to prevent LOC-I.

Proposed LOC-I Methodology for UAVs

LOC-I definitions and supporting frameworks for analysis of events have been devised for commercial aviation (Bromfield & Landry, 2019) and general aviation (Smith & Bromfield, 2022). However, for UAV LOC-I events, a considerably different operating environment, requires a more flexible, qualitative approach. The Accident Route Matrix (NASEM, 1998) offers a more suitable 'hybrid' approach benefitting from the high-level fixed categories of Human Factors Analysis and Classification System (HFACS, 2014) in combination with the flexibility of AcciMap framework (Systems Thinking Lab, 2023). This approach has enabled the inclusion of all identified (and normalised) UAV accident causal and contributory factors. The use of a timeline in the ARM approach also allows these factors to be presented in a sequence of events leading to LOC-I, including operator's response, the main post-flight procedures, and recovery response whether successful or not (Figure 1).

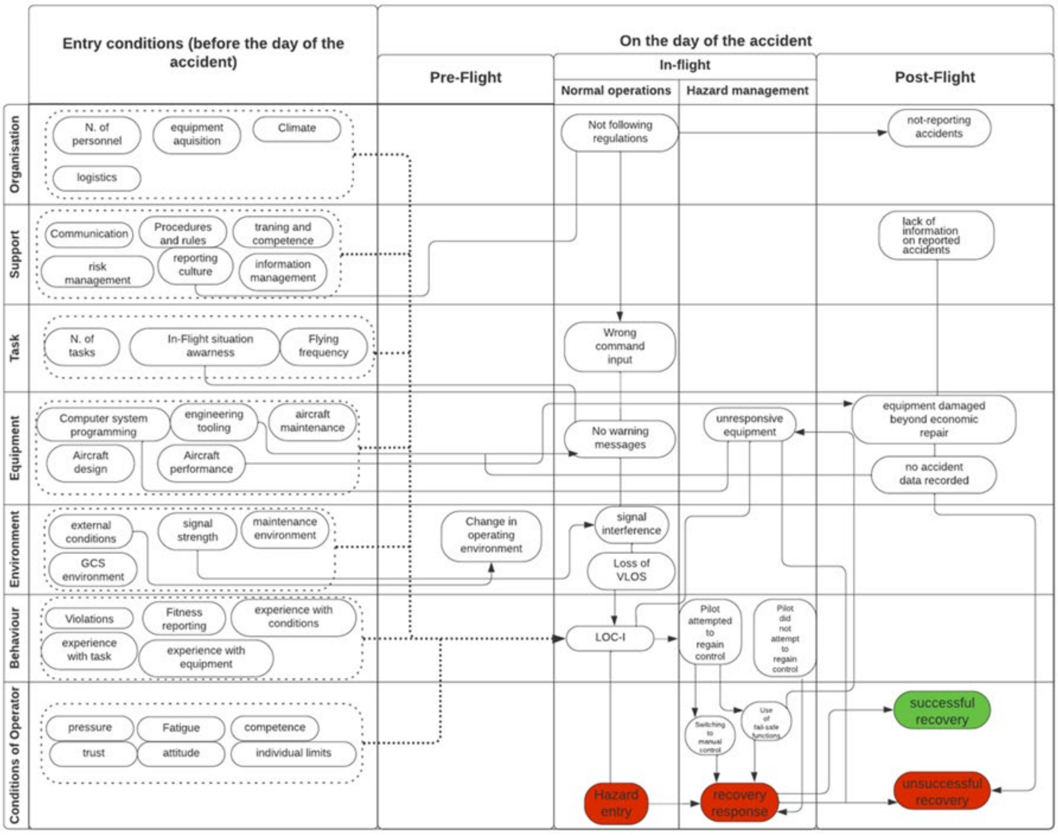


Figure 1: UAV LOC-I framework (adapted from ARM)

References

Bromfield, M. A. & Landry. S. J., (2019) "Loss of Control In Flight (LOC-I) – Time to Re-define?", (AIAA 2019-3612), 2019 Aviation Technology, Integration, and Operations Conference, Dallas, Texas, USA, 17-21 June 2019 https://doi.org/10.2514/6.2019-3612

HFACS Inc, (2014), The HFACS Framework, HFACS Incorporate, [retrieved 17 Feb 2023]. Available from: https://www.hfacs.com/hfacs-framework.html

IATA (2019), Loss of Control In-Flight Accident Analysis Report Edition 2019 Guidance Material and Best Practices. 2019.

McCarley, J.S. & Wickens, C.D., (2004), Human Factors Implications of UAVs in the National Airspace, Federal Aviation Administration.

NASEM (1998), Improving the Continued Airworthiness of Civil Aircraft: A Strategy for the FAA's Aircraft Certification Service. National Academies of Sciences, Engineering, and Medicine, Washington, DC: The National Academies Press. https://doi.org/10.17226/6265

Nilsson, S.J. (2011), Relationship between Recent Flight Experience and Pilot Error General Aviation Accidents, PhD Dissertation, Northcentral University, Prescott Valley, Arizona, USA, May 2011.

Smith, J. & Bromfield, M.A., (2022) "General Aviation Loss of Control in Flight Accidents: Causal and Contributory Factors, " Journal of Air Transportation, Vol. 30, No. 4 (2022), pp. 137-153 https://doi.org//10.2514/1.D0286

Systems Thinking Lab, AcciMap, [retrieve 17 Feb 2023]. Available from: https://systemsthinkinglab.com/accimap/

Weedon S., Cutler V & Revell S., (2014), A Trend Analysis of Human Factors Issues in UK Military Aviation, Royal Airforce Centre of Medicine, Bedfordshire, UK.

Design induced non-compliance: influences on pedestrian and cyclist behaviour at level crossings

Gemma J. M. Read[1,2], Nicole Liddell[1], Pia Sauer[1], Paul M. Salmon[1]

[1]Centre for Human Factors and Sociotechnical Systems, University of the Sunshine Coast, Queensland, Australia [2]School of Health, University of the Sunshine Coast, Queensland, Australia

SUMMARY

Collisions at rail level crossings remain a pressing concern, with the influences on user behaviour a critical area of research. This paper reports the findings of an observational study of pedestrian and cyclist non-compliant behaviours at 10 rail level crossing sites in Australia. The findings illustrate the diversity in crossing designs and how these differences may influence behaviour. General recommendations are provided, alongside the need to consider context-specific risk controls.

KEYWORDS

Rail level crossings, Pedestrians, Cyclists, Risk, Non-compliant behaviour

Introduction

Rail level crossings (RLXs) pose a safety threat to road users in Australia, with 39 level crossing collisions involving a person or road vehicle and 725 near miss incidents since 2016 (ONRSR 2021). The key non-compliant behaviour of entering the RLX when gates/booms are closed, known as "bypassing", increases the risk of collision between trains and road users, including pedestrians and cyclists. However, the factors underpinning this behaviour are not clear. The aims of this study were to: (1) Improve our understanding of bypassing behaviour of pedestrians and cyclists at RLXs; and (2) Identify factors that influence bypassing behaviour and risk at RLXs.

Method

Data were collected using the Behavioural Assessment Tool for Rail Level Crossings (BAT-RLX; Read & Salmon, 2016). The BAT-RLX coding scheme was applied to analyse road user behaviour at 10 sites in Victoria, Australia. For each site, eight hours of video footage was analysed, generally over two weekdays, during peak periods (7:15 to 9:15 am and 3:30 to 5:30 pm). Coding was supported by the Noldus Observer XT software package.

Results & Discussion

A total of 201 bypasses were identified during the study period. The majority of bypassers were male (62%), were adults (91%), and the majority of bypass events involved a single user as opposed to bypassing in a group (68%). Just over half of bypassers did not check for trains prior to bypassing (55%) and the majority did not engage with technology (e.g., use a mobile phone / device or wear headphones) while bypassing (93%).

For each site, there was generally a clearly preferred method of bypassing (either via the pedestrian gates or via the road boom barriers), aligned to the differences in the physical design of the site, including desire lines. Most bypass events (58.2%) occurred during the morning peak period; however, this trend was more pronounced for some sites with adjacent train stations, indicating this behaviour may be attributed to time pressure associated with catching a train to reach work / school on time. However, at other sites with adjacent stations, less clear differences between the peak periods were observed suggesting that behaviours are driven by other factors at these sites. Where the destination of the bypasser could be determined at RLXs with an adjacent train station, just under half (48.39%) proceeded to the train station versus other destinations after bypassing. For some sites, a clear majority of bypassers were observed to continue to the railway station; however, at other sites there was a more even spread of those going to the station versus another destination. This suggests that while catching a train may be an influencing factor in some cases, it does not account for all bypasses at sites with an adjacent railway station.

Figure 1 shows the wait times of pedestrians and cyclists before bypassing. There is large variation in wait times, indicating potential contextual influences. However, over half of the sites had a median wait time of less than one minute before users bypassed, indicating a low tolerance to wait.

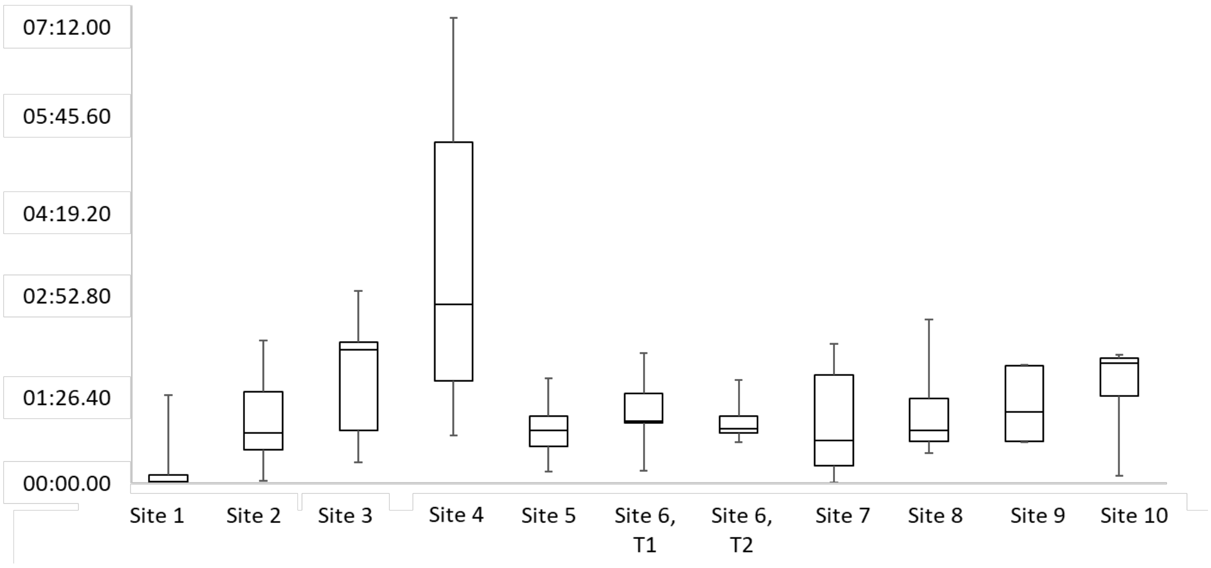


Figure 1: Wait times prior to bypassing

Conclusion

The findings provide in-depth data regarding the demographics and circumstances of bypass events involving pedestrians and cyclists at RLXs. They also highlight the diversity in RLX designs and how these may influence behaviour (i.e. desire lines) as well as the need to explore how to manage the conflict between user wait time tolerance and actual waits in a busy metro rail environment. General recommendations to improve safety include improving user information, improving design to guide users to preferred paths, and providing stronger physical barriers for pedestrians and cyclists. However, the findings regarding the influence of context mean that broad assumptions should not be used when assessing risks and determining risk controls for individual sites. Site-specific data collection and risk assessment are important inputs to RLX risk management.

Acknowledgements

This research is funded by the Victorian Railway Crossing Safety Steering Committee via Metro Trains. We thank Vanessa Beanland, Tony Carden, Leannda Read, Kerri Salmon and Kaytee Faulkner for their contributions to data collection.

References

Office of the National Road Safety Regulator (ONRSR). (2021). *Rail Safety Report 20-21*. Available from: https://www.onrsr.com.au/publications/corporate-publications/rail-safety-report

Read, G., & Salmon, P. (2016). *Development of a Level Crossing Behavioural Assessment Tool. Report for the Victorian Railway Crossing Safety Steering Committee (VRCSSC)*. University of the Sunshine Coast, Centre for Human Factors and Sociotechnical Systems.

Design Blindspots: User testing clinical IT systems

Lauren J Morgan[1,2] & Paula Pryce[2]

[1] Morgan Human Systems Ltd, [2] Shrewsbury & Telford NHS Trust, UK

SUMMARY

In early 2021, the MHRA launched its guidance on applying human factors and usability engineering to medical devices including drug-device combination products in Great Britain (MHRA, 2021). In its guidance it states: A usability engineering process can, and should, be applied by device manufacturers in the identification, assessment and mitigation of potential patient and user safety risks; also in the analysis of incidents that have occurred, in order to identify learning and put into place corrective actions to improve device design

However, experience in hospital healthcare is that many devices and IT systems are often poorly designed and continue to contribute to patient safety risks. A seminar Harvard Business Review paper stated: "to fix physician burnout, we must first fix the electronic patient record". In everyday work in our hospitals we see examples where poor device and IT design is making clinicians lives harder, and decreasing patient safety as a consequence. We have a workforce cataclysm, of which the state of hospital devices and IT is possibly contributing to rather than helping to fix. We explore a simplified multi-method approach to user testing to identify patient safety and usability risks. We present evaluations of 3 clinical IT systems, showing how user testing conducted correctly easily identifies these safety risks. The MHRA guidance as currently stands is not being used fully by suppliers, we need to consider how to strengthen its impact.

KEYWORDS

Maternity, Clinical, IT, design, usability

Introduction

In healthcare, the model most widely accepted to describe the system of work is the Systems Engineering in Patient Safety (SEIPS) model (Carayon et al., 2006). In that model, one of the 6 main components is tools and technology, which signifies the importance of considering the impact of tools and technologies on patient outcomes. This is a long acknowledged fact in human factors/ergonomics (Grandjean, 1988).

In healthcare, there is a fraught relationship with technology. The National Programme for IT (NPfIT) was one of the largest publicly funded IT project failure. Most people outside healthcare would reasonably be surprised at the lack of digitisation, and lack of design in the technology that is available. Many had been calling for a significant time for engagement of HF in the design of clinical devices (Waterson, 2014). In Jan 2019, the much awaited guidance from the MHRA was launched, largely based on the guidance from the US Food and Drug Administration (FDA) (FDA, 2016). This guidance requires manufacturers of devices to follow a usability engineering process. What is considered a device in healthcare is complex, some IT systems are included, some are not. However, there is nothing preventing those not strictly covered by the guidance from taking its advice.

Informal conversations with manufacturers shows that they believe meeting the guidance is easy to fake. That they have had to change very little in their processes. We sought to explore what the reality of this is through the testing of some devices in the healthcare setting. Could we identify risks to patient safety that could, and should have been picked up if the suppliers had followed a usability engineering process.

Methods

We tested different devices and IT systems currently being implemented/offered for sale to the Trust. We used a mixed methods approach, based on the ISO 9241 standard for usability. For satisfaction we used the system usability scale, a validated tool. We also present an updated version for use in healthcare.

Results

The testing revealed that even the devices and systems felt to be the most useful, still had usability issues. Perhaps most alarmingly, during testing significant risks to patient safety were easily identified. Users verbalised “this is the first time anyone has allowed us to talk about IT as if it is making us less safe, which we all know, but can’t say”.

Discussion

The impact of poor design of clinical devices and IT systems is vast in terms of patient safety outcomes and staff experience. This will only increase as the push for paperless healthcare increases. We must put pressure on designers of these systems to fully embrace the MHRA guidance and deliver real benefits to healthcare as a result.

References

Carayon, P.A.S.H., Hundt, A.S., Karsh, B.T., Gurses, A.P., Alvarado, C.J., Smith, M. and Brennan, P.F., 2006. Work system design for patient safety: the SEIPS model. *BMJ Quality & Safety*, *15*(suppl 1), pp.i50-i58

Grandjean, 1988. Fitting the task to the Man.

Food and Drug Administration, 2016. Applying Human Factors and Usability Engineering to Medical Devices: Guidance for Industry and Staff.

Waterson, P 2014. Health information technology and sociotechnical systems: A progress report on recent developments within the UK National Health Service (NHS), Applied Ergonomics,

Forging Links between Safety Critical Task Analysis and Incident Investigation

James Bunn, Simon Dunford, Neil Hunter & Mary Marshall

Health and Safety Executive, Energy Division, UK

SUMMARY

Safety Critical Task Analysis (SCTA) is an established methodology in high-hazard industries for identifying and managing the human contribution to the risk of major accidents. It seeks to identify which tasks could contribute to the initiation or escalation of a major accident, and then identifies the steps within those tasks where additional controls may be required to mitigate against the effects of human failure. An incident investigation that incorporates examination of human and organisational factors would also be expected to examine where and how failures occurred, the efficacy of control measures in place and where remedial actions are required.

This paper describes how the incident investigation process and investigation findings and the SCTA process can support each other and sets out a case for connection and improved alignment between them. This is a connection that has been underexplored to date, yet which has benefits for both risk management and human factors integration.

KEYWORDS

SCTA, Task Analysis, Investigation, Risk Management

Introduction

Investigations provide valuable insights that guide improvements in risk management and control, albeit after the incident and its associated damage have occurred. This has supported risk management methodologies that can look ahead and proactively identify where and how failures may occur. Within design and engineering, Hazard Identification (HAZID) and Hazard and Operability (HAZOP) studies provide a structured way to achieve this and are established elements of major accident risk management. Safety Critical Task Analysis (SCTA) provides a multi-phase, structured approach to analysing and assessing non-technical aspects. SCTA shares features of HAZID/HAZOP such as the involvement of experienced people to provide input to the assessment, use of guidewords to support discussion and a qualitative approach. SCTA focuses on human activities and seeks to identify where and how human failure could occur within tasks that could contribute to the initiation or escalation of a major accident.

Investigation and SCTA share common elements. Both rely on the construction of a sequence of actions and events within which failures can be identified. Both seek to understand how failures occur, the control measures in place to prevent the failure or mitigate the effects, and any resulting gaps. Both seek to identify improvement actions. This provides potential for alignment between the two processes, and a potential information input and feedback mechanism to support them both. Examples are shown in Figure 1.

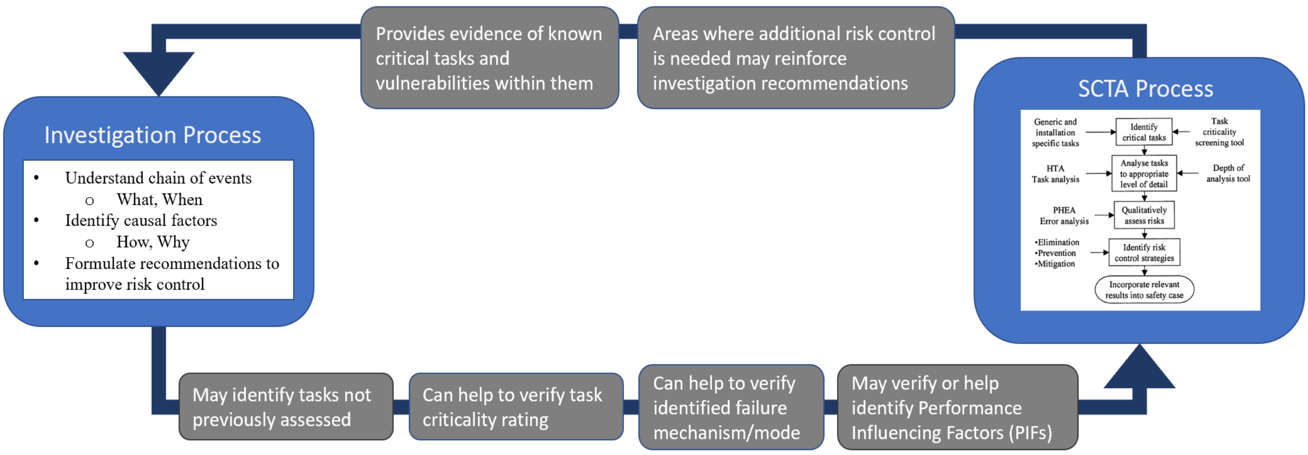


Figure 1: Potential alignment and feedback link between SCTA and investigation processes.

Evidence from Inspections, SCTA projects and Investigations

Evidence from focused Human Factors inspection of UK offshore duty holders in 2021-2022 and from investigation and SCTA work for UK onshore duty holders in 2018-2022 revealed no links between investigation and SCTA processes of the type illustrated in Figure 1. For the duty holders and organisations involved, the processes were functionally distinct. Specifically:

- No investigation procedures that were reviewed as part of the inspections identified SCTA data as a potential evidence source or included the SCTA process as part of post-investigation, follow-on activities.
- No investigation procedures that were reviewed specified that the task(s) that were involved in the incident should be identified.
- No SCTA procedures that were reviewed included using investigation information to inform or verify the analyses of identified tasks.

A review of existing SCTA guidance found that investigation findings were briefly included as a potential source of information for constructing an inventory of tasks for safety criticality screening. The Energy Institute guidance (2020) describes how SCTA is a proactive way to manage risk, unlike investigations which are reactive. It also states that SCTA is used by some high-hazard companies as the established approach to managing the human component of MAH risk. No links or acknowledgment of a potential operational relationship between SCTA and the investigation processes is included. A review of existing Human Factors in Investigation guidance (CIEHF 2020, HSE 2005, Energy Institute 2008) found no link between investigation and critical task analysis.

Discussion

The duty holders who were inspected were generally at an early stage of the SCTA process and had not reached a level of maturity with the process by which they could begin to explore operational interactions with other processes. Similarly, some (though not all) of the duty holders were at a relatively early stage of Human Factors integration with their investigation process and were focused on developing HF competence and adopting a structured HF analysis methodology. Existing guidance on SCTA (HSE 1999, Energy Institute 2020) generally focuses on the methodology and the activities required for the various phases of the process. The existing guidance does not explore how the process might link to other areas of risk or operational safety management. Taken together, there is no clear pointer in the existing guidance that might prompt a duty holder or organisation to link the processes, other than through their own initiative.

Forging Operational Links

This paper does not present a new, distinct methodology. Rather, it sets out a case for improved alignment between SCTA and investigation processes to support operational risk management. The following examples build on the information provided in Figure 1. They do not provide an exhaustive list but help to illustrate how operational links between investigations and SCTA might work in practice by describing activities in the present tense in a hypothetical organisation that has established SCTA and investigation processes.

Investigation input to SCTA

- The organisation's procedure or governing document for SCTA specifies that the incident recording database is interrogated when developing an inventory of tasks to screen for safety criticality. The focus is principally on high potential incidents. The personnel responsible for carrying out the SCTA work review the results and note the work activities that were involved in the incidents, and any indications of human failure or Performance Influencing Factors (PIFs) that were included in the investigation reports. Recommendations from the investigation(s), particularly recommendations that have not been completed are also noted.
- The work activities identified from the investigation information are compared with the list of tasks compiled from lists of procedures or work instructions, input from operators and the other sources of information used for defining the various types of tasks performed at the site / installation / functional area. Where there is an existing procedure, work instruction or similar the investigation information is included as notes. Where no such documentation exists, the investigation information is used to describe distinct tasks that are included in the inventory and taken forward for screening.
- During screening, where a task has been associated with an incident, or if the incident still has recommended actions outstanding, the screening output is considered alongside this information. If the screening indicates a low level of safety criticality the option is provided to revise the priority of the task for further analysis, based on the investigation information.
- Where a task has been prioritised for further analysis and has been associated with an investigated incident, the notes about human error, PIFs and recommendations made from the incident recording database are fed into the task analysis and the Human Error Analysis (HEA) work phases. This information informs the discussion that the analyst(s) have with operators when tasks are observed, talked-through and analysed in detail.
- Recommendations for additional control measures arising from the Human Error Analysis are checked against the recommendations noted from the investigation information. An investigation recommendation that is closed-out but which reappears from the HEA work prompts re-examination of the issue. An investigation recommendation that has not been closed-out and which results from the HEA work is flagged and given a higher priority rating.

These activities acknowledge a risk of cognitive bias, particularly where the analyst(s) and workshop participants' thinking may be overly influenced by investigation information and does not consider alternative failure modes, consequences or requirements for additional control measures. This can be mitigated against through awareness raising in the education process for SCTA facilitators, and by specifying in the SCTA procedure that investigation information shall not dictate the course of the HEA work.

SCTA input to investigations

- The investigation procedure specifies that the task(s) involved with the incident shall be identified as part of the investigation terms of reference for describing the incident and sequence of events (the "what happened" part that is common to investigations).
- The investigation team checks the inventory of tasks for the site/installation or functional area that was created to provide input to safety critical task screening. The team records a) any matches and b) any potential omissions. This is done once the investigation team have sufficient knowledge of the incident to determine what tasks might have been involved, for example when the main fieldwork phase concludes and the data analysis phase commences.
- Where a task from the inventory has been identified as safety critical and a HEA has been carried out, the information is used in the investigation analysis process. The HEA provides information about PIFs, potential human failure modes, existing risk controls that were cited at the time of the analysis and where additional risk control requirements were identified. These elements are all of relevance to the investigation and might be corroborated through the investigation work or highlight areas where risk control was inadequate.
- Where a previously unidentified task is revealed through the investigation, it is added to the task inventory and screened for safety criticality. The actual and potential severity of the incident is considered during screening and prioritisation for subsequent analysis as described in the previous subsection. The investigation information is used to inform the analysis, also as described in the previous subsection.
- Where a known task has been involved in the incident, an action arising from the investigation is to review the SCTA work associated with that task. This may include re-screening the task for safety criticality and adjusting the previous results, reviewing the task analysis and reviewing the HEA as required.
- Recommendations arising from the investigation are compared with any additional risk controls identified from the HEA process for involved tasks and are reviewed or re-prioritised in the same way as described in the previous subsection.

It is important that this aspect of the investigation-SCTA link is not misinterpreted as "correcting the homework" of the SCTA analyst(s), rather that the link helps to validate the proactive work and provides opportunities to potentially improve the quality of both the proactive and reactive work.

The examples provided above demonstrate that, where an organisation has an investigation process and a SCTA process in place, no major changes to either process are required. Instead, the operational links can be achieved through formal acknowledgement of the respective processes in the relevant procedures or governing documentation and through specifying actions within them. This provides the formal mandate for the actions and helps to ensure that they will be carried out and followed-up. By modifying existing processes rather than introducing a new process, the workload and resource requirements are not anticipated to be burdensome, and this would be expected to become more efficient as the organisation's experience with working with SCTA and human factors in investigations matures.

Conclusion

Forging links between investigation and SCTA means that retrospective findings from investigations can support proactive findings from SCTA work and vice versa. This supports the quality of both processes. It is particularly useful for helping to identify more complicated failure modes or Performance Influencing Factors (PIFs) at key task steps, which can be challenging even for experienced workshop participants. It also helps to identify safety critical tasks that are but might not have been previously identified. Forging links between SCTA and investigation reinforces the concept of SCTA as a dynamic process that reacts to changes and new information, like any other risk assessment. Linking investigated incidents from an organisation's incident recording database to tasks in the SCT register can help an organisation to identify 'hot spots' in particular task types or functional areas, which in turns supports prioritisation of resources for risk management. Using investigation information in the HEA phase of SCTA (rather than just as input when constructing a task inventory) helps to support learning from incidents, specifically through capturing relevant information in a proactive risk management process and reviewing actions required for improved risk control.

The findings described in this paper indicate that guidance in both areas could be developed further, particularly as human factors in investigations and SCTA become more established operational applications of Human Factors in high-hazard industries and operational links become more apparent. Some duty holders in the petroleum industry have linked SCTA to Management of Change (MoC) processes and have used SCTA to support focused risk management projects. This indicates that beyond the SCTA / Investigation link presented in this paper, SCTA has further unrealised potential for connection with risk management processes and as a supporting framework for human factors integration.

References

Chartered Institute of Ergonomics and Human Factors (2020). Learning from Adverse Events. CIEHF White Paper. Published via www.ergonomics.org.uk/resource/learning-from-adverse-events

Energy Institute (2020). Guidance on Human Factors Safety Critical Task Analysis. Second Edition. Energy Institute, London.

Energy Institute (2008) Guidance on investigating and analysing Human and Organisational aspects of Incidents and Accidents. Energy Institute, London.

Health and Safety Executive (2000). Human Factors Assessment of Safety Critical Tasks. Offshore Technology Report OTO 1999/092.

Health and Safety Executive (2005). Investigating accidents and incidents. HSG245. HSE Books.

A Human Factors review of "the Blue Puffer" asthma reliever inhaler

Deborah Stratford[1] & Susan Whalley-Lloyd[1,2]

[1]School of Health, Science and Wellbeing, Staffordshire University, UK, [2]Lloyds Risk Services Ltd, UK

SUMMARY

The literature reports that Asthma Inhaler technique has not improved during the last forty years, despite improvement strategies focused on educating users to improve their technique and compliance. This is particularly critical for reliever inhaler users when 'use error' may result in a full asthma attack and possible death. This paper presents a pilot study Human Factors design review of the standard UK reliever inhaler, commonly referred to as 'the blue puffer'. The results indicate a mismatch between 'work as done' and 'work as imagined' and that this mismatch appears to be influenced by the design of the inhaler. Conceptually it appears possible to improve the design of technical components of the inhaler system to reduce use errors and hence improve patient safety. This would require appropriate scenario and user testing, with any changes being integrated into the system as a whole.

KEYWORDS

Patient Safety, Inhaler Design, Human Factors, Human-Centered Design, Sociotechnical systems

Introduction

Asthma related deaths have increased in the UK over the last decade with over 1,400 deaths from asthma attacks in 2018 alone, including a 42% increase in people aged 35 to 44 years (Asthma and Lung UK, 2019).

Self-administered drug delivery through inhalation is the primary method for treatment of respiratory diseases such as asthma and chronic obstructive pulmonary disease (COPD) (Schreiber 2020). This is via Preventer inhalers. In addition to using a preventer, most of these patients will be prescribed a Reliever inhaler to use during an acute episode of breathing difficulty or when they have signs of an asthma attack. Reliever inhalers can also be prescribed for short term use for those with temporary breathing difficulties. In the UK the current most common asthma reliever is the Ventolin Evohaler pressurized metered-dose inhaler (PMDI), this is blue in colour and commonly referred to as 'the blue puffer'.

Inhaler technique has not improved during the last forty years despite improvement strategies. These strategies have largely focused on improving human action by providing additional education to improve technique and compliance (Sanchis et al, 2016). Furthermore, it has been shown that as few as 9% of healthcare professionals tasked with educating patients in inhaler use had adequate knowledge of all prescribed steps. Healthcare professionals were unable to demonstrate the correct technique required to ensure effective medication dose delivery (Baverstock et al, 2010).

Aim

The aim of this study was to identify aspects of the reliever inhaler system that appear to impact user performance and to consider Human Factor improvement opportunities with the potential to reduce ‘use error’ and improve optimal drug dose delivery.

Method

This pilot study was conducted between January and April 2022 by a cohort of Human Factors postgraduate students from Staffordshire University, all currently working within the healthcare system.

The overall method was based on the Human Centered Design Process (Gilero, 2022). The first four steps were used to provide a framework for reviewing the reliever inhaler sociotechnical meso system, see Figure 1. The system comprised of; inhaler users, medical professionals, the inhaler device plus spacer, instruction leaflet, packaging and an educational video provided by Asthma and Lung UK (Asthma and Lung UK, 2021)

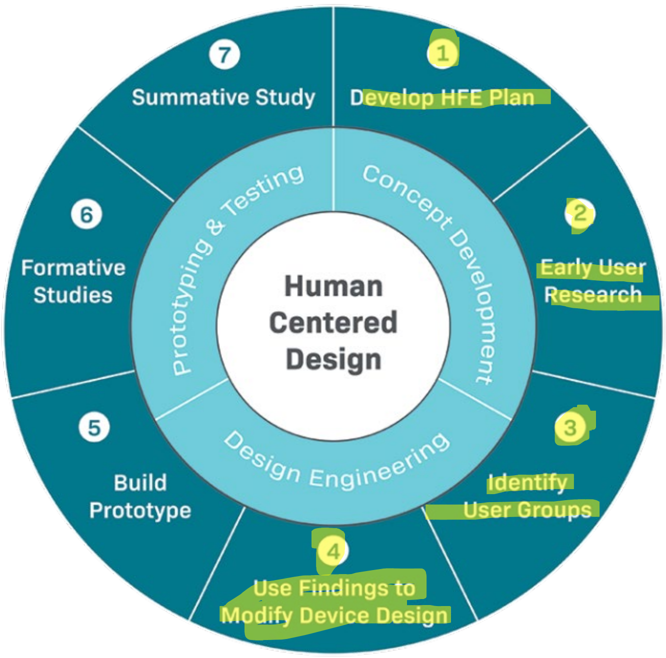


Figure 1: The Human Centered Design process

Phase One: Develop HFE Plan

Step 1: Select the Reliever inhaler for review

An initial literature review identified use problems associated with reliever inhalers and which Reliever inhaler was predominantly used within the UK. A photograph of the Blue Puffer inhaler and Spacer is provided in Figure 2. Note that Spacers are often recommended for use with a reliever inhaler to help ensure the full dose of medication reaches the lungs.

Step 2: Compile Study Brief and key dates

The brief provided an overview of patient safety issues associated with Reliever inhalers, photographs of the components, links to manufacturer patient instruction sheets and links to the patient training videos available on the Asthma and Lung UK website (Asthma and Lung UK, 2021). The pilot study ran from 5th March – 8th April 2022

Figure 2: The Ventolin, Blue Puffer Inhaler and associated Spacer (photographs by C Saunders, 2022)

Phase Two: Early User Research undertaken by Student Group

Step 1: Full group brainstorming session

An initial live virtual group brainstorming session was conducted. The purpose was to facilitate a shared understanding of the reliever inhaler system and to prompt consideration of the range of users, their capabilities and limitations and why these may impact effective system interaction.

Step 2: Interviews and Observations

Two separate sessions were conducted live virtually. Each Interview session included Task Observation and walk-through-talk-through with a reliever inhaler user. Each participant had used a reliever and preventer inhaler for more than twenty years. One participant was male aged 55 and one female aged over 60.

Phase Three: Identify User Groups and their needs

Each reviewer was responsible for identifying likely users based on the initial brainstorming session, user interviews, task analysis and their own personal research.

Phase Four: Use findings to modify device designs

Phase four comprised the Human Factors review and suggestions for Design Modifications. To support Phase Four, the Inclusive Design Wheel was used as a prompt, see Figure 3 (unknown author). This covers the basic human sciences applied during design.

The review of each technical element was based on the concept of 'use error' to focus attention on any mismatch between design and user capabilities. It also recognised that poor design can create 'stressed' users who then have an increased potential for error (Wears et al, 2016). Conversely user-centred design can minimise training needs (Wears et al, 2016).

Figure 3: The Inclusive Design Wheel

Results

Figure 4 presents the results of the initial Brain Storming Activity which was used to help understand the different elements of the system from a micro and meso system level. This is followed by an example of the tasks as imagined and as done.

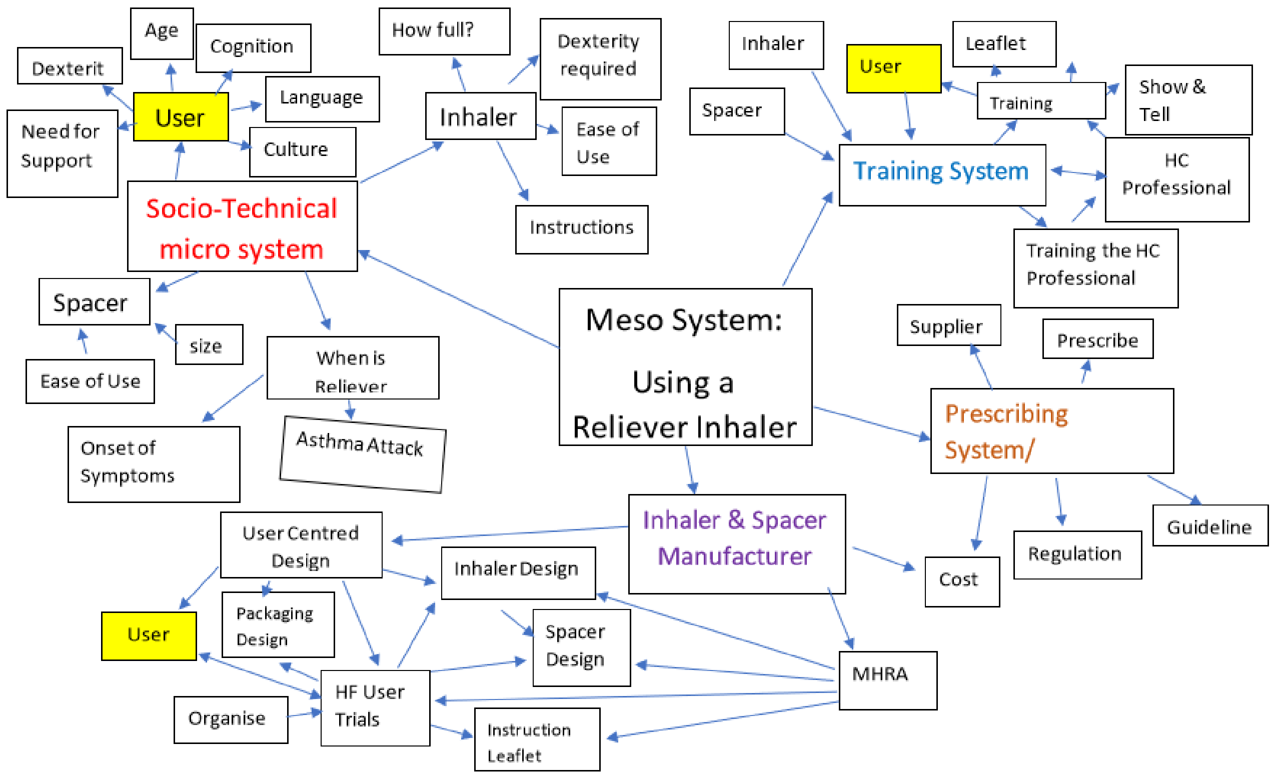


Figure 4: Reliever Inhaler Sociotechnical System Network

In terms of the specific results, this paper focuses on the reliever inhaler and its Patient Instructions Leaflet.

Instructions for using the Ventolin reliever inhaler

Figure 5 provides an example of generic user instructions provided by a Ventolin Inhaler supplier-Work as Imagined (https://www.ukmeds.co.uk/ventolin).

> In order to use your Ventolin inhaler correctly you should follow this step-by-step guide:
>
> 1. Take the cap off the mouthpiece.
> 2. Hold the inhaler upright and shake it very well.
> 3. Exhale and tilt your head a bit.
> 4. Place the mouthpiece in your mouth and enclose your lips on the mouthpiece.
> 5. Breathe in slowly and deeply with your mouth, press down the canister firmly and let out one puff of Ventolin. Continue inhaling slowly.
> 6. Take the inhaler off your mouth; hold your breath for about five seconds. Exhale with your nose.
> 7. Place the cap back.
>
> Should you need to take another puff, wait around 60 seconds before you repeat steps 2-7 above.

Figure 5: User Instructions

This example does not match the user video instructions ((Asthma and Lung UK, 2021), the available instruction leaflets (PILS) within inhaler packaging, nor the tasks identified during the user task observations, Figure 6. This demonstrates some of the problems encountered by front-line healthcare professionals who first instruct the inhaler user and the inhaler users themselves as they try to determine how to use their inhaler. Recall of instructions is in any case an issue over time without use and in stressful situations e.g. experiencing the start of an asthma attack or difficulty with breathing.

1. Inhaler grasped in palm of hand using fist grip
2. Other hand removes the dust cap from mouth piece
3. User Shakes inhaler up and down in vertical position twice
4. Mouthpiece placed in mouth between lips (no head tilt)
5. Top of the inhaler canister is depressed with thumb whilst slowly breathing in
6. Inhaler is immediately re-pressed and user breaths in again due to insufficient dose
7. Dust cap not replaced on mouth piece (this increase the risk of mouthpiece damage, contamination and foreign objects becoming lodged in the mouthpiece preventing its use

Figure 6: Observed Task Sequence

Specific Results for the Inhaler Device

Anthropometric Considerations

The Users in the pilot study did not use the recommended pinch grip. One reason may be that the length of the reliever inhaler at 85mm is difficult for most users to comfortably and securely hold in a pinch grip. The lack of contact points between the inhaler and the hand may also be an issue. The observation sessions showed that users grasped the inhaler in their fist (grasp grip) and used the thumb to activate it – rather than their index finger.

Biomechanical interaction

Device contours were suboptimal with raised edges and corners creating tissue hotspots which Pheasant and Haslegrave (2006) recommend should be eliminated. The radial shape of the PMDI was not a consistent cylindrical or elliptical shape which prevented even points of contact with the anatomy of the human hand (see Figure 7). This also prevented an optimal compression grasp grip (Pheasant and Haslegrave, 2006; Matuszek and Drobina 2018) which was the users grip of choice.

The device cross sectional diameter was 25mm at its widest point and 15mm at its narrowest point, which was below an “optimal diameter of 30-50mm” as suggested by Pheasant and Haslegrave (2006 pp.153) for optimal grip strength. A ridged zone was located on the bottom of the Reliever inhaler to aid a pinch grip, see Figure 8, which did not reflect Work as Done (WAD) and therefore, both users were unaware of this design feature.

The surface material was smooth plastic which may reduce grip when interacting with the viscoelasticity and lubrication of human skin (Pheasant and Haslegrave, 2006).

The observed user's shoulder and arm position, necessary to comply with the prescribed head tilt (see the Asthma and Lung UK (2021) video), involved a significant forward arm and shoulder flexion this reduces effective usability for users with comorbidities and upper limb injuries.

Figure 7: Radial Contours + Hands

Figure 8: Ridged Area

Figure 9: Inhaler differences

Sensory Interactions

The shape of the reliever inhaler matches those that are perceived to be associated with a typical grasp grip.

Visually, the reliever inhaler was the exact size and shape as the preventer inhaler, the only differences were colour plus a 'V' shaped raised area detectable visually and by touch on some Ventolin reliever inhalers, see Figure 9. There were no visual or physical markings on the preventor inhaler device. Note: the mouthpiece protection cap for the preventor inhaler in this photograph has been lost.

Specific Results for Patient Information Leaflet (PIL)

The PIL (Cardill, 2022) revealed a Fog Index readability score of 10.89. A score of 7-8 is considered ideal, whereas above 12 is too complex (EDUTAS, 2019). No alternate language options were provided. Copies of the leaflet are available in Braille, large print or audio but need to be requested by phoning a designated number and this opportunity may be missed by the user.

Physically the Patient Instruction Leaflet (PIL) was difficult to open as the leaflet was tightly folded and paper was thin which meant it could tear easily (Cardill 2022). In any case, Bix (2016) reported that written patient information has a negligible impact on user adherence and that leaflets may be lost or discarded. Both participants stated that they threw away the PIL without reading it.

Discussion and Recommendations

Cognitive Considerations

The affordance priming effect (APE) of the reliever inhaler design appears to have been a potential reason for the identified gap between work as imagined and work as done in terms of user grip choice, fist rather than pinch grip. The affordance priming effect (APE) states that the visual stimuli of design and shape significantly influences the neural to physical pathway mechanisms and subsequent physical hand and arm interactions with an object (Makris et al 2013). This indicated a possible ambiguity in device design (Natraj et al, 2013) it looks to the user as though a fist grip would be the best choice.

In terms of task steps the differences between the Patient Insert Leaflet supplied with the inhaler and the basic instructions provided on the internet demonstrate inconsistencies in work as imagined. Given this situation, there can be no surprise that actual work as done does not comply. The users within this pilot study did not read their instruction leaflets and did not know that training videos existed. They were therefore reliant on Health care workers to train them and they then had to remember what they were shown. The Literature suggests that older people with airway diseases have a high prevalence of cognitive deficits such as reduced information processing, memory, attention and concentration levels which can result in poor competence and compliance in inhaler techniques (Jin Song et al, 2022). Regardless of cognitive defects, Long Term memory and recall degrade over time, which places additional importance on good design rather than training.

Physical Considerations

Dexterity and grip can be impacted by room temperature and humidity with hand skin temperature being the most significant factor; grip strength has been shown to decrease by 16% after two minutes of environmental exposure to 5ºC (Prasetyo, 2020; Vincent and Tipton, 1988). Given that one trigger of breathing problems can be moving into cold temperatures, this may increase the likelihood of needing to use a reliever inhaler in these conditions. Age and gender also impact physical ability and dexterity and the average grip strength declines in both men and women (Viana et al, 2007). The range of shoulder mobility has been found to significantly decline with age, including active flexion and abduction movements with external rotation ranges particularly declining in females (Gill et al, 2020) this is important given that the inhaler is expected to be used with a recommended backwards head tilt.

Suggested integrated System changes

Any changes to one element of the system needs to be followed through into the rest of the system technical components in order to achieve a robust system that supports the user. For example: A QR code could be included on the Inhaler and PIL that takes the user to an Instruction Video (Kenyon, 2022); An inhaler design that supports the grasp grip should be shown on PIL infographics and the instruction video. Combined concept design adaptations need to be user centered and incorporate Human Factors theory. User trials should consider how to mimic actual use scenarios to appreciate the impact of stress on use errors and to ensure that the design can cope with the actual conditions a user is faced with.

Conclusions

The current design of the PMDI device inhaler system provides inadequate support to typical users, such as those involved within this pilot study. Users had to make significant adaptations from techniques that they were not even aware of, despite regular asthma reviews with healthcare staff. Both users made identical adaptations.

There are numerous resource disincentives for healthcare providers to invest in human centred design (Bix, 2016; Wears et al, 2016). It is therefore important to assess whether a change in one or more elements of the Reliever Inhaler meso socio-technical system is cost beneficial. To do this it is necessary to consider the whole patient journey. At the start, improved technical design is likely to reduce the amount of patient training required, reducing health carer contact time. In addition, an improved design that reduces use error should offer lifecycle cost benefits including a reduction in ambulance journeys, hospital inpatient stays and ultimately a reduction in patient deaths. It is inappropriate to concentrate on the unit price of the inhaler device itself without placing this in context with a full cost benefit analysis.

References

Age UK (2016). The Internet and Older People in the UK – Key Statistics

Allen, S., Jain, M., Ragab, S. et al (2003). Acquisition and short-term retention of inhaler techniques require intact executive function in elderly subjects. *Age and Ageing* Vol 32:3

Almomani, B., Mokhemer, E., Al-Sawalha N. et al (2018). A Novel Approach of Using Educational Pharmaceutical Pictogram for Improving Inhaler Techniques in Patients with Asthma. *Respiratory Medicine* 143:103-108.

Asthma and Lung UK (2019). Asthma death toll in England and Wales is the highest this decade Available at: https://www.asthma.org.uk/about/media/news/press-release-asthma-death-toll-in-england-and-wales-is-the-highest-this-decade/ [Accessed online March 2022]

Asthma and Lung UK (2021). How to use a PMDI inhaler. Available at: https://www.asthma.org.uk/advice/inhaler-videos/pmdi/#externalvideoblock8123 [Accessed online March 2022]

Ayres, J. (2022). *A Human Factors Design Review: Element 3*. Staffordshire University, available at: https://acrobat.adobe.com/id/urn:aaid:sc:EU:4fee6837-315d-4760-ba15-94739d031a6d

Baverstock, M., Woodhall, N., Maarman V. et al (2010). Do Healthcare Professionals Have Sufficient Knowledge of Inhaler Techniques in Order to Educate Their Patients Effectively in Their Use? *Thorax,* December 2010 Vol 65 Suppl 4. DOI:10.1136/thx.2010.150979.45

Bix, L. (2016). The Human Factor – Pharmaceutical Packaging: Packaging Materials Containers and containment Services. *Being Human.* Available at: https://www.ns-healthcare.com/analysis/the-human-factor-pharmaceutical-packaging-4879328/# [Accessed online March 2022]

BSI Group (2015). British Standard 62366-1:2015+A1:2020 Medical Devices: Application of Usability Engineering to Medical Devices. Available at: https://bsol-bsigroup-com.ezproxy.staffs.ac.uk [Accessed online March 2022]

BSI Group (2021). British Standard 15223-1:2021 Medical Devices: Symbols to be Used with Information to be Supplied by the Manufacturer. Available at: https://bsol-bsigroup-com.ezproxy.staffs.ac.uk [Accessed online March 2022]

Budiu, R. (2014). Memory Recognition and Recall in User Interfaces. Neilson Norman Group. Available at: https://www.nngroup.com/articles/regocnition-and-recall/ [Accessed online March 2022]

Cardill, L. (2022). A Human Factors Design Review: packaging and patient information leaflet. Staffordshire University, available at: https://acrobat.adobe.com/id/urn:aaid:sc:EU:d5fe3d52-1ff7-459b-95b9-0629de6a2ec3

Carrington, A. (2022). *Assessing Asthma UK Website's Training Videos for the Use of Pressurised Metered Dose Inhaler.* Staffordshire University, available at: https://acrobat.adobe.com/id/urn:aaid:sc:EU:754b9855-ba1a-4c84-b0ac-23526e0cae51

CIEHF (2022). Chartered Institute of Ergonomics and Human Factors, *Guidance to Help Design Effective Useable Work Procedures for Health and Social Care Teams*. Available at: https://www.ergonomics.org.UK [Accessed online February 2022]

Dowse, R., and Ehlers, M. (2001). The Evaluation of Pharmaceutical Pictograms in a Low-literate South African population. *Patient Education and Counselling* Vol 45:87-99

EDUTAS (2019). The Fog Index. University Outreach EDUTAS The University of Oaklahoma available at: https://outreach.ou.edu/educational-services/education/edutas/comp-center-

landing-page/knowledgebases/program-evaluation-knowledgebases/task-1-write-report/fog-index/ [Accessed online March 2022]
Equality Act (2010). Available at: https://legislation.gov.uk/ukpga/2010/15/contents
Forskett, K. (2022). A Human Factors Design Review. Staffordshire University, available at: https://acrobat.adobe.com/id/urn:aaid:sc:EU:894560b0-70d2-4d05-867c-af084e9b230e
Fountain, N. (2022). *A Human Factors Design Review*. Staffordshire University, available at: https://acrobat.adobe.com/id/urn:aaid:sc:EU:037e160b-3aff-4acd-a27d-53b8b12dd865
Gianlorenco, A., and Braccialli, L. (2010). Influence of the object texture related to the grip in individuals with Down Syndrome. *Fisioterapia em Movimento*, Vol23(2): 229-238 DOI: 10.1590/5013.51502010000200007
Gill, T., Shanahan, E., Tucker G. et al (2020). Shoulder range of movement in the general population: age and gender stratified normative data using a community-based cohort. *BMC Musculoskeletal Disorders* 2020 21:676.
Gillero (2022). The Importance of Human Factors for Medical & Drug Delivery Devices Available at: https://www.gilero.com/news/importance-human-factors-medical-drug-delivery-devices/#:~:text=Human%20factors%20studies%20examine%20how,in%20relation%20to%20the%20product. [Accessed online March 20222]
GOV.UK (2020). Guidance Medicines: packaging, labelling and patient information leaflets. Available at: https://www.gov.uk/guidance/medicines-packaging-labelling-and-patient-information-leaflets [Accessed online March 2022]
GSK- GlaxoSmithKline UK (2021). Packing Leaflet: Information for the User. Ventolin Evohaler 100 micrograms salbutamol sulphate. Available at: https://medicines.org.uk/emc/product/850/pil#gref [Accessed online March 2022]
Guastello, S. (2014). *Human Factors Engineering and Ergonomics*: A Systems Approach (2nd edition) pp.103-109. CRC Press Taylor Francis Group ISBN: 13-978-1-4665-6009-3
IMSN-International Medication Safety Network (2013). Position Statement: Making Medicines Naming, Labeling and Packaging Safer. Available at: https://www.intmedsafe.net/contents/AboutIMSN.aspx [Accessed via Staffordshire University Online March 2022]
Jin Song, M., Yee Kim S., Ae Kang, Y. et al (2022). The relationship between cognitive function and competence in inhaler technique in older adults with airway disease. *Geriatric Nursing* 43 pp.15-20
Jowett, V. (2022). *Ventolin instructions and packaging: A Human Factors Review:* Staffordshire University, available at: https://acrobat.adobe.com/id/urn:aaid:sc:EU:a6e0e37d-e4d6-44a3-af9e-a73d781c2b2e
Kenyon, S. (2022). *Element 3: Instructional Video*. Staffordshire University, Available at: https://acrobat.adobe.com/id/urn:aaid:sc:EU:3efea8eb-49d9-4bfa-9b52-5d8b58622ea8
Klonoff, D., Bassock, S., Dwyer A. et al (2020). Evaluating the usability and safety of the semaglutide single-dose pen-injectors through summative (Human Factors) usability testing. *Journal of Diabetes Investigation*, Vol 12(6): 978-987
Kovoor, J., McIntyre, D., Chick, W. et al (2021). Clinician-Created Educational Video Resources for Shared Decision-making in the Outpatient Management of Chronic Disease: Development of an Evaluation Study. *Journal of Medical Internet Research* Vol 23(10): e26732 pp.1-13 DOI: 10.2196/26732
Lindh, A., Theander, K., Arne, M. et al (2019). Errors in inhaler use related to devices and to inhalation technique among patients with chronic obstructive pulmonary disease in primary health care. *Nursing Open*, Vol 6:1519-1527. DOI: 10.1002/nop2.357

Lotlikar, A. (2022). *Access to the Video*. Staffordshire University, available at: https://acrobat.adobe.com/id/urn:aaid:sc:EU:bfbf4867-ebb9-4b90-82bd-c6c4b159122d

Makris, S., Grant, S., Hadar, A. et al (2013). Binocular vision enhances a rapidly evolving affordance priming effect: Behavioural and TMS evidence. *Brain and Cognition* Vol 83 pp.279-287

Matuszek, J., and Drobina, R. (2018). Designing Handles of Hand Tools in the Aspect of Comfort and Safety: Design for Accessibility. *Ergonomics For People with Disabilities*. ISBN: 9783110617832

McNab, D., McKay, J., Shorrock, S. et al (2020). Development and application of 'systems thinking' principles for quality improvement. *BMJ Open Quality* 2020;9:e000714, DOI:10.1136/bmjoq-2019-000714

MHRA-Medicines and Healthcare Products Regulatory Agency (2020). *Best practice guidance on the labelling and packaging of medicines.* Available at: https://assests.publishing.service.gov.uk

Ministry of Housing Communities & Local Government (2019). The English Indices of Deprivation 2019 (IoD2019)

Montagne, M. (2013). Pharmaceutical pictograms: A model for development and testing for comprehension and utility. *Research in Social Administrative Pharmacy Vol* 9(5) pp.609-620 DOI: https://doi.org/10.1016/j.sapharm.2013.04.003

NASA- National Aeronautics and Space Administration (2019). *NASA TLX Task Load Index Paper and Pencil Version*. Available at: https://humansystems.arc.nasa.gov/groups/tlx/downloads/TLXScale.pdf

Natraj, N., Poole, V., Mizelle, J. et al (2013). Context and hand posture modulate the neural dynamics of tool-object perception *Neuropsycholgia* Vol 51 pp.506-519.

ONS- Office for National Statistics (2021). *Internet users*, UK:2020

Otunbade, E. (2022). *Human Factors Inhaler and Spacer Design Review: a case for change.* Staffordshire University, available at: https://acrobat.adobe.com/id/urn:aaid:sc:EU:ff4c5145-5ffe-4fb1-95c3-82189f2b9655

Pheasant, S., and Haslegrave, C. (2006). *Bodyspace: Anthropometry, Ergonomics and the Design of Work* (3rd edition), CRC Press, Taylor Francis Group, ISBN: 10.0-415-28520-8

Prasetyo, Y. (2020) Factors Affecting Gross Manual Dexterity: A structural Equation Modeling Approach. In *Proceedings of the 2020 2nd International Conference on Management Science and Industrial Engineering (MSIE 2020).* 304-308.

Rosen, M., Salas, E., Tannenbaum, S. et al (2012). Simulation-Based Training for Teams in Health Care: Designing Scenarios, Measuring Performance, and Providing Feedback, in Carayan P (ed.) *Handbook of Human Factors and Ergonomics in Health Care and Patient Safety* (2nd edition) pp.573-591, CRC Press Taylor Francis Group.

Sanchis, J., Gich, I., and Pederson, S. (2016). *Systematic Review of errors in Inhaler Use: Has Patient Technique Improved Over Time?* Available at: https://dx.doi.org/10.1016/j.chest.2016.03.041 [Accessed online March 2022]

Schreiber, J., Sonnenburg, T., and Luecke, E. (2020). Inhaler devices in asthma and COPD patients – a prospective cross-sectional study on inhaler preferences and error rates. *BMC Pulmonary Medicine* 20:222

Shipway, B. (2022). *Human Factors Design Review*. Staffordshire University, available at: https://acrobat.adobe.com/id/urn:aaid:sc:EU:76284485-df8c-4e79-944a-78bc35689cbb

Simpson, R., Knowles, E., and O'Cathain, A. (2020). Health Literacy Levels of British Adults: a cross-sectional survey using two domains of the Health Literacy Questionnaire (HLQ). *BMC*

Public Health Vol 20:1819 Available at: https://doi.org/10.1186/s12889-020-09727-w [Accessed online March 2022]

Świeczkowski, D., and Kułacz, S. (2021). The use of the Gunning Fog Index to evaluate the readability of Polish and English drug leaflets in the context of Health Literacy challenges in Medical Linguistics: An exploratory study. In *Cardiology Journal 2021, Vol. 28, No. 4, 627–631 DOI: 10.5603/CJ.a2020.0142*

Tillman, B., Fitts, D., Wesley, E. et al (2016). *Human Factors and Ergonomics Design Handbook*, (3rd edition), McGraw Hill Education, ISBN: 978-0-07-170287-4

Travis, D., and Hodgson, P. (2019). *Think Like a UX Researcher*: How to Observe Users, Influence Design, and Shape Business Strategy, CRC Press, Taylor Francis Group ISBN: 13.978-1-138-36529-2

Tretinjak, M. (2015). The Implementation of QR Codes in the Educational Process. *MIPRO* 2015, 25-29 May 2015. pp.833-835 Available at: https://ieeexplore-ieee-org.ezproxy.staffs.ac.uk/stamp/stamp.jsp?tp=&arnumber=7160387

Vianna, L. C., Oliveira, R. B., and Araujo, C. G. (2007). Age-Related Decline in Handgrip Strength Differs According to Gender. *Journal of strength and conditioning research* ISSN:1064-8011 Volume: 21 Issue:4 Page: 1310-1314 DOI:10.1519/00124278-200711000-00058 Available at: https://www-proquest-com.ezproxy.staffs.ac.uk/docview/213061686?OpenUrlRefId=info:xri/sid:summon&accountid=17254

Vincent, M., and Tipton, M. (1988). The effects of cold immersion and hand protection on grip strength, *Journal of Aviation Space Environment Med* Vol 59(8): 738-41 PMID: 3178622. Available at: https://www.pubmed.ncib.nlm.nih.gov [Accessed online March 2022]

Wears, L., Rollin, J., and Fairbanks, T. (2016). Design Trumps Training. Pediatrics/Editorial, *Annals of Emergency Medicine*. DOI: 10.1016/j.annemergmed.2015.10.014. Available at: https://www-sciencedirect-com.ezproxy.staffs.ac.uk/science/article/pii/S0196064415013761?via%3Dihub [Accessed online March 2022]

WHO- World Health Organisation (2021). *Asthma.* Available at: https://who.int/news-room/fact-sheets/detail/asthma [Accessed online March 2022]

Augmented Design with Voice Recognition and Auditory Alerts in the Flight Deck

Niall Miranda, Jean-Baptiste Bonotaux & Wen-Chin Li

Safety and Human Factors in Aviation MSc, SATM, Cranfield University, United Kingdom

SUMMARY

Devices using voice recognition and verbal auditory alerts have advanced rapidly in the modern world. In an era that drives the world using voice commands, aviation has lagged to implement this technology in cockpits. Contrary to the rapid pace of advancements in cockpit instruments, auditory alerts have also been largely primitive. With such technologies gaining rapid acceptance in many industries, it is about time to consider adopting the same in cockpits, especially to support pilots during enhanced operational workload. This paper conducts a use case study on the prevailing aural technologies in the cockpit whilst exploring the augmentation of contemporary technologies in voice recognition and verbal auditory alerts to aid pilots' cognition and reduce mental workload.

KEYWORDS

Augmented Reality, Cognitive Support, Semantic Alert, Situation Awareness, Voice Recognition

Introduction

From adopting glass cockpits to substituting cockpit crew with avionics, cockpit technologies have evolved rapidly benefitting aviation safety and enhancing the piloting experience. However, the usage of auditory alerts has largely been primitive whilst the use of voice recognition is practically non-existent in civil aviation. The use of navigation systems with verbal auditory alerts in the automotive industry has helped drivers experience enhanced cognition and situation awareness whilst reducing mental workload. During a dynamically changing environment, it is pivotal for pilots not to lose situation awareness whilst constantly shuffling to perceive information from the cockpit displays and the outside environment. The judicious use of verbal auditory alerts and voice recognition in tandem with artificial intelligence can positively impact cockpit dynamics and improve cognition leading to enhanced flight safety (Lin et al., 2022).

Augmenting Voice Recognition

Voice recognition in cockpit purposes to serve pilots' commands to be interpretable by the flight computers. This technology has made its way into the military cockpits of Eurofighter (Smith, 1999) and Lockheed Martin F-35 (Schutte, 2007) to facilitate pilots to conduct operations in highly dynamic environments by alleviating operational workload. Speech synthesis technology is employed in operating autopilot modes and aircraft configuration whilst also performing cockpit functions such as operating buttons, levers, and switches supplementary to manual inputs. However, voice recognition has lagged to commence in civil aviation. With the advancements of technology in avionics, the complexity of the human-machine interface keeps escalating the operational perplexity for pilots leading to a higher mental workload and loss of situation awareness.

Acoustic alerts versus synthesized speech

Acoustic alerts have been tolerably functioning in cockpits as transient auditory signals to alert pilots of impending dangers or adverse changes in the aircraft's configuration and performance. These sounds do not however impart an intuitive link between the alert and its target function thereby imposing an initial learning curve to comprehend the aural taxonomy. Synthesized speech, unlike acoustic sounds, is nimble to be assimilated. However, scant synthesized speech alerts are in existence. Considering the multitude of alerts required by the pilots to maintain situational awareness, it is ergonomically congruent to instate synthesized verbal alerts to keep the pilots informed of the various modes and states of the aircraft systems. Additionally, whilst synthesized speech is combined with acoustic alerts to deliver critical information, a shorter time is required for cognition (Kearney et al., 2016).

Use Case

Human-computer interaction systems using augmented reality such as Microsoft HoloLens with voice command can aid in reducing physical demand and cognitive load whilst increasing situation awareness (Li et al., 2022). Supplementary to voice commands, gaze commands could also be implemented owing to its resistance to unintentional activations, and the minimized workload and reduced time required to accomplish the required action (Isomoto et al., 2020). Augmenting visual cues in tandem with aural annunciations to propagate the required information can prevent inadvertent operations and misconfiguration of the aircraft (Conner et al., 2012). Implementation of the HoloLens in a cockpit environment (see fig. 1) can be devised to safeguard checklists during pilot performance degradation triggered by enhanced operational workload and a dynamically changing environment. This use case demonstrates that such technologies could improve the pilot's safety and efficiency during crucial phases of flight by ensuring acceptable levels of performance without omission of actions. The holographic features of the HoloLens in combination with voice commands, gaze commands, and synthesized speech can optimize the efficiency and safety of the flight by:

- Providing a step-by-step list that is predominantly on the pilot's line of sight
- Displaying components requiring action when a checklist task is enunciated
- Ensuring that the correct action has taken place in accordance with the checklist
- Automatically switching over to the next task on the checklist

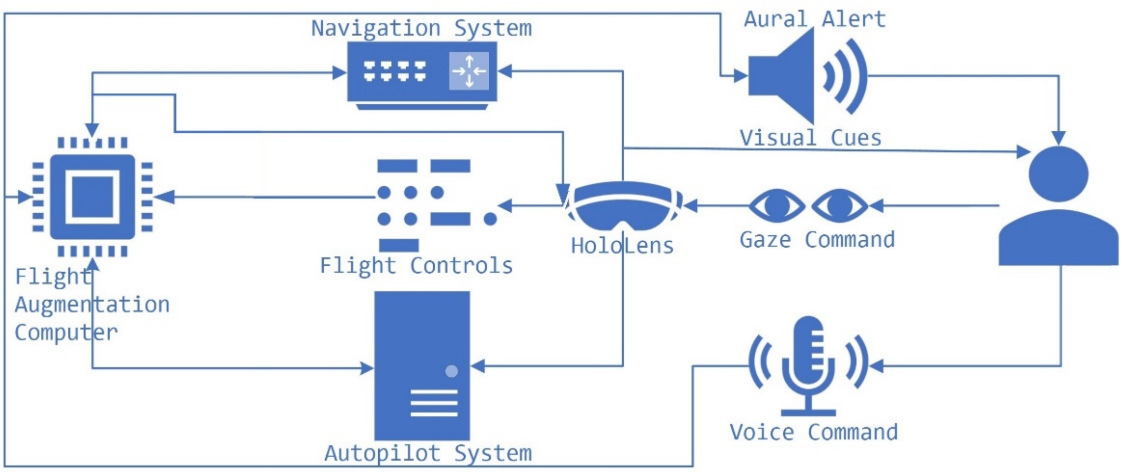


Figure 1. Use case of voice command and gaze command with an augmented reality device

Conclusion

Successful implementation of augmented support to pilots either voice commands or gaze-command will be a potential solution for the future single-pilot flight deck. Furthermore, the augmented reality device has to be able to recognize variations in speech prosody and speech from a comprehensive vocabulary which can reduce the pilot's mental workload. The saliency of visual and auditory messages on both inputs and outputs in the flight deck requires further investigation to support single-pilot operations in the future.

References

Conner, K. J., Feyereisen, T., Morgan, J., & Bateman, D. (2012). Cockpit displays and annunciation to help reduce the loss of control (LOC) or lack of control (LAC) accident risks. *AIAA Guidance, Navigation, and Control Conference 2012*. https://doi.org/10.2514/6.2012-4763

Isomoto, T., Yamanaka, S., & Shizuki, B. (2020). *Gaze-based Command Activation Technique Robust Against Unintentional Activation using Dwell-then-Gesture*. In Proceedings of the 46th Graphics Interface Conference (Graphics Interface 2020), CHCCS, Tronto, Canada, May 28-29, 2020 https://www.iplab.cs.tsukuba.ac.jp/paper/international/isomoto_gi2020.pdf

Kearney, P., Li, W. C., & Lin, J. J. H. (2016). The impact of alerting design on air traffic controllers' response to conflict detection and resolution. *International Journal of Industrial Ergonomics*, *56*, 51–58. https://doi.org/10.1016/j.ergon.2016.09.002

Li, W. C., Zhang, J., Court, S., Kearney, P., & Braithwaite, G. (2022). The influence of augmented reality interaction design on Pilot's perceived workload and situation awareness. *International Journal of Industrial Ergonomics*, *92*. https://doi.org/10.1016/j.ergon.2022.103382

Lin, Y., Ruan, M., Cai, K., Li, D., Zeng, Z., Li, F., & Yang, B. (2022). Identifying and managing risks of AI-driven operations: A case study of automatic speech recognition for improving air traffic safety. *Chinese Journal of Aeronautics*. https://doi.org/10.1016/j.cja.2022.08.020

Schutte, J. (2007, October 11). *Researchers fine-tune F-35 pilot-aircraft speech system*. Human Effectiveness Directorate. https://www.afmc.af.mil/News/Article-Display/Article/154961/researchers-fine-tune-f-35-pilot-aircraft-speech-system

Smith, C. J. (1999). Design of the Eurofighter human machine interface. *Air & Space Europe*, *1*(3), 54–59. https://doi.org/10.1016/S1290-0958(00)88429-4

Take me home, country road: Comparative optimism, mind-wandering in an automated simulator

Rachael A. Wynne, Angus McKerral, Sophie Withers & Kristen M. Pammer

School of Psychological Sciences, The University of Newcastle, Australia

ABSTRACT

Mind wandering while driving has been shown to factor in distracted driving, a critical cause of road crashes in Australia. With the implementation of autonomous vehicles onto the road network proposed as occurring in the near future, lies the potential for increased mind wandering, as cognitive engagement in driving is lessened. Part of the potential appeal of such vehicles is the ability to perform non-driving related tasks while in an automated driving mode. This study presents an analysis of drivers' subjective experience of two prolonged drives in a driving simulator set to automated driving. Half of the participants were permitted to engage in non-driving tasks, to simulate potential future features. Participants provided summaries of their experience and preparedness to take-over control following two critical events. This study explores the themes of participants' subjective experience and how this relates to mind-wandering, comparative optimism of driving behaviour, and readiness to respond to take-over events.

KEYWORDS

Automated vehicles, comparative optimism, non-driving related tasks, subjective experience, driving simulator

Introduction

Advancements in artificial intelligence, machine learning, and automation technology over the last 30 years as resulted in substantial changes to the technology landscape and by extension the operation of automotive vehicles (de Winter et al., 2014; Merat et al., 2014; Meyer et al., 2017). The release of autonomous vehicles into the road environment has been highly promoted and accompanied with claims of allowing the driver to engage in other non-driving related tasks such as using their mobile phones, working, or even sleeping. Level 3-4 autonomous vehicles require the driver to maintain a degree of environment maintenance for the duration of the drive and be physically and cognitively prepared to take over the if necessary.

The engagement of 'Autopilot' modes reduces the driver's physical and cognitive responsibilities (Endsley, 2017). Such features have the potential to significantly reduce the number of motor vehicle crashes and fatalities by supporting the driver with pre-emptive warnings. However, if the driver is not equipped to take over from the vehicle when prompted, the outcomes have been shown to be serious and even fatal.

This presents an inherent contradiction, with the driver being in a position to respond to the driving environment compared to being 'freed-up' to engage with other tasks. While requiring drivers to remain focused and attentive towards the road, while not driving the vehicle, introduces human factors risks like fatigue, reduced situation awareness, and increased distractibility. These factors have been demonstrated to impact perception, attention and decision-making ability while driving (Guo, et al., 2016).

While physical distractions such as mobile phone use, entertainment systems, and eating have been studied extensively in research, inattention or distraction due to mind wandering does not feature heavily in the road safety literature. Mind wandering occurs without the present of an external stimulus, and is defined as when conscious mental focus deviates to matters unrelated to the task at hand (Smallwood et al., 2003, Smallwood & Schooler, 2006). Research by Burdett and colleagues (2016; 2019) has explored this occurrence in everyday manual driving. Indicating its high prevalence, prevailing even when aware of the phenomena. Mind wandering was found to be more common in unfamiliar environments, when fatigued and when on longer drives (Burdett et al., 2016; 2016). With the advances associated with automated vehicles seen to be a benefit to issues such as fatigued, this would potentially increase the opportunity for mind wandering if sufficient cognitive load is not maintained.

Comparative optimism encompasses the notion that individuals consider themselves to be more skilful and less susceptible to risk than the average person of the same age (Gosselin et al., 2010; Harré & Sibley, 2007; Harris & Middleton, 1994; Shepperd et al., 2002). Comparative optimism is pervasive across age groups, and it is considered that this would extend to participants' perceived ability to monitor an autonomous vehicle.

This paper explores the themes of the participants' subjective experience of an automated vehicle across two different roadways (rural and city); and between subjects' comparison of the engagement of a non-driving related task. The paper considers how this relates to their experience of mind-wandering, comparative optimism of driving behaviour, and their preparedness to two take-over events.

Method

Forty-four participants, (25 female, 19 male) aged between 18 – 36 years (M = 23.52, SD = 4.84) were recruited the University population. Although external advertising was used, all participants were undergraduate psychology students, receiving research awareness credit points as part of their involvement. Potential participants were screened for susceptibility to simulator sickness. Participants held either a provisional (n = 17) or open driver licence (n = 27), and therefore were able to drive on their own. All participants drove at least once a week, with most driving occurring on suburban roads (40-50km/hr) and urban roads (60km/hr).

The study incorporated a 2 x 2 mixed design. Roadway was a within-subjects variable including city and rural roadways. The engagement with non-driving related tasks was a between subjects variable, with participants randomly, and evenly allocated to allowing the use of non-driving related tasks while in automation mode, or not. The experiment was part of a larger study testing

This study used a STISIM M300WS Driving Simulator at The University of Newcastle, Callaghan Campus, running a beta version of STISIM Drive 3. The automation software allowed drivers to experience highly automated driving modes by initiating self-driving mode in which the automation offers adaptive cruise control, monitoring both longitudinal and lateral movements. The driving simulator is considered to be medium fidelity (Wynne et al., 2019); encompassing multiple screens and a broad field of view, "arcade" vehicle controls with pedals, seat, and steering wheel, as well as a motion platform. An image of the visual environment and the simulator appear as Figure 1 and 2 respectively.

Figure 1: Example of Typical Road Scape Scene During Experiment

Figure 2: Visual Depiction of Simulator Configuration.

Note. Not visible, motion base and arcade seat.

Prior to the testing session, all participations completed a basic demographics and driving experience questionnaire. Upon arrival at the session and introduction, participants were given a verbal brief regarding study instructions and the capacity of the simulator's self-driving mode. Participants experienced two practice trials approximating five minutes each; the first to familiarise themselves with the physical simulator and the displayed road scape with the second to familiarise themselves with self-driving mode and how to initialise and disengage it safely. During this trial, participants were issued a takeover request from the system during a non-critical point. Those participants assigned to the non-driving related task condition were advised they could use which ever devices or materials they brough with them to the testing session when the vehicle was in automated driving mode.

The main trials ran for approximately 50 minutes each and were counterbalanced (rural or city). Following a prolonged period of automated driving, self-driving mode was automatically disengaged during a critical event, and participants were required to take over control of the vehicle. The take-over-requests consisted of a visual and auditory prompt instructing participants to takeover manual control. One critical event consisted of posted roadworks with workers appearing around heavy machinery. The other event where self-driving mode was disengaged consisted of a car pulled over flashing hazard lights, fallen boxes across the left-hand lane and pedestrians over on the other side of the road. After their main trial had concluded, participants completed an interview on their subjective experience, topics of mind-wandering and what strategies were used to maintain an alert state. Participants were also asked of their preparedness to take back control of the vehicle if, and when, required. The timeline of events appears as Figure 3.

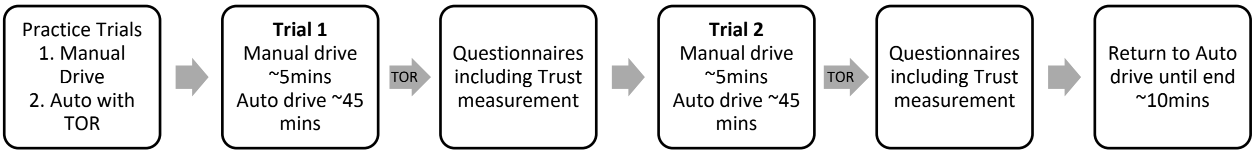


Figure 3: Timeline of Events for Testing Session.

Note. Sessions were counterbalanced for roadway (rural versus city).

Preliminary Results and Discussion

Preliminary analysis of 30 participants of equal non-driving related task conditions has been performed, with the remaining analysis to be conducted imminently. From these analyses, mind wandering was found in both conditions. The content of the mind-wandering, however, was more varied when non-driving related tasks were not permitted. At an individual level, participants reported having varied their thoughts more over the time, and one indicated it was easy to drift off with no engagement. Those able to use non-driving related tasks felt they were able to maintain engagement with the driving task, and the content of their mind wandering was focused on the task they became engaged with (e.g., phone call or social media).

The preliminary results suggest some mental engagement in the driving task was maintained while the use of a non-driving related task, however the impact on driving behaviour and time taken to respond to this request will need to be explored.

References

Burdett, B. R., Charlton, S. G., & Starkey, N. J. (2016). Not all minds wander equally: The influence of traits, states and road environment factors on self-reported mind wandering during everyday driving. *Accident Analysis & Prevention, 95*, 1-7.

Burdett, B. R., Charlton, S. G., & Starkey, N. J. (2019). Mind wandering during everyday driving: An on-road study. *Accident Analysis & Prevention, 122*, 76-84.

Endsley, M. R. (2017). Autonomous driving systems: A preliminary naturalistic study of the Tesla Model S. *Journal of Cognitive Engineering and Decision Making, 11*(3), 225-238.

Harré, N., & Sibley, C. G. (2007). Explicit and implicit self-enhancement biases in drivers and their relationship to driving violations and crash-risk optimism. *Accident Analysis & Prevention, 39*(6), 1155-1161.

Harris, P., & Middleton, W. (1994). The illusion of control and optimism about health: On being less at risk but no more in control than others. *British Journal of Social Psychology, 33(*4), 369-386.

Gosselin, D., Gagnon, S., Stinchcombe, A., & Joanisse, M. (2010). Comparative optimism among drivers: An intergenerational portrait. *Accident Analysis & Prevention, 42*(2), 734-740.

Guo, M., Li, S., Wang, L., Chai, M., Chen, F., & Wei, Y. (2016). Research on the relationship between reaction ability and mental state for online assessment of driving fatigue. *International Journal of Environmental Research and Public Health, 13*(12), 1174-1186.

Merat, N., Jameson, H. A., Lai, F. C., Daly, M., & Carsten, O. M. (2014). Transition to manual: Driver Behaviour when resuming control from a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour, 27*, 274-282.

Meyer, J., Becker, H., Bosch, P. M., & Axhausen, K. W. (2017). Autonomous Vehicles: The next jump in accessibilities? *Research in Transportation Economics, 62*, 80-91.

Shepperd, J. A., Carroll, P., Grace, J., & Terry, M. (2002). Exploring the causes of comparative optimism. *Psychologica Belgica, 42*(1/2), 65-98.
de Winter, J.C., Happee, R., Martens, M.H., & Stanton, N.A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation Research Part F: Traffic Psychology and Behaviour, 27*, 196-217.
Wynne, R. A., Beanland, V., & Salmon, P. M. (2019). Systematic review of driving simulator validation studies. *Safety Science, 117*, 138-151.

Non-technical skills in recreational environments – the case of Mountain Guides

Amy Irwin[1], James Thacker[2] Julie Evans[1] & Gabriel Brame[1]

[1]University of Aberdeen, UK. [2]Mountain Assurance, UK.

SUMMARY

Mountaineering is a potentially high-risk activity, encompassing uncertain and dynamic terrains and adverse weather conditions. Within this context mountain guides work with client groups to plan activities and enhance safety. Despite the integral nature of mountain guides, and their responsibility for other individuals, to date there has been no research examining the non-technical skills necessary for safe and effective performance within this role. The current study investigates the non-technical skills utilised by mountain guides and evaluates potential influencing factors that might impact performance. Semi-structured interviews were conducted with 19 British Mountain Guides. Content analysis led to the development of six non-technical skill categories (situation awareness, decision-making, leadership, teamwork & communication, cognitive readiness, task management) with 18 associated elements. Identified influencing factors included fatigue, tunnel vision, competitiveness, task pressure and client behaviour. The results highlight the importance of these skills within the mountain guide context and suggest a need for non-technical skills training going forward within this complex and high-risk domain.

KEYWORDS

Mountain guides, non-technical skills, situation awareness

Introduction

Guided mountaineering is often considered an aspect of adventure tourism, where participants engage in strenuous activity during both summer and winter seasons, most often hiking, skiing and rock climbing, within a natural mountain landscape (Rebelo et al., 2018). The level of activity can vary from potentially high risk (e.g. rock climbing) through to lower risk (e.g. hiking through lower mountain ranges). Client ability level and experience can also range from novice to experienced. Regardless, the mountain guide is integral to these expeditions, from planning the activity through to managing client wellbeing and safety (Rebelo et al., 2018). Accredited mountain guides are experienced mountaineers who have undertaken training and assessment to secure the recognised IFMGA carnet to become a British Mountain Guide (BMG). The guide has responsibility for the safety and wellbeing of their clients during any activities undertaken and will usually also engage in an element of education and training with their group (Rebelo et al., 2018). The combination of mountaineering expertise requirements, alongside client management, has led previous researchers to suggest that the role of mountain guide is unique in its juxtaposition of performance, safety and service within a dynamic and uncertain environment (Girard, Caroly & Falzon, 2020).

Non-technical skills

Despite the high-risk and complex nature of mountain guiding, there is a lack of research evaluating the range of non-technical skills that might be vital for this role. The primary focus of previous research has been on decision-making due to the potentially serious, or life-threatening consequences, of poor decisions in this context (Walker & Latosuo, 2016; Stewart-Patterson, 2004). Non-technical skills (NTS) are defined as the social (e.g. teamwork, leadership, communication) and cognitive (e.g. situation awareness, decision-making, cognitive readiness, task management) skills that, in addition to technical knowledge, enable safe and effective performance (Flin, O'Connor & Crichton, 2017). These skills have been studied across a range of industries, including aviation (Flin et al., 2018), offshore drilling (Sneddon et al., 2006), healthcare (Reader et al., 2006), maritime operations (Fjeld et al., 2018) and agriculture (Irwin et al., 2022). Work in these industries has identified that lapses in NTS can be a key cause of accidents and injuries, for example, failures in teamwork have been linked to adverse events during surgery (Catchpole et al., 2008). The adverse consequences are not limited to the workers themselves but can also encompass patients or clients. However, to our knowledge, no research has as yet focused on these skills within a recreational or service driven context – where clients are highly involved in the production or development of activities, including goal setting (Girard et al., 2020). By expanding NTS research into a new domain, and one with recreational elements rather than a purely work-based focus, we can enhance our understanding of the application of these skills in different contexts and potentially develop new categories and elements tailored to this domain. As such, the current study seeks to add to the NTS literature through an exploration of the NTS utilised by British Mountain Guides.

In addition, it is important to consider NTS with a systems lens, evaluating how individuals' function within a complex system, including both non-technical influencing factors (e.g. time pressure) and NTS. Understanding the inter-relationship between skills and influencing factors enables greater understanding of both adverse incidents, and the elements that support work, or in this case guiding, proceeding as expected (Naweed & Murphy, 2022). Previous research highlights that within the context of mountain guiding, the client themselves can present a risk, via pressure to achieve a goal, risk taking behaviours, lack of experience or failure to abide by the guides decisions (Girard et al., 2022). This presents a new and unique consideration for the utilisation of NTS, whereby the 'team' consists primarily of fee-paying clients, as opposed to co-workers.

Study aims

The focus of the current study was on the identification, and description, of key NTS categories, elements and associated behaviours necessary for safe mountain guiding. In addition, to understand the utilisation of these skills within the mountain guide context, including barriers and facilitators, we also aimed to describe key factors that might influence NTS performance and safety.

Method

Participants

A total of 19 Mountain Guides were recruited via direct email and Mountain Guide mailing lists. All participants were members of the British Association of Mountain Guides and had a minimum of two years as a mountain guide post-qualification.

Interviews

Participants took part in semi-structured interviews via the telephone which were a maximum of one hour in duration. Each interview consisted of five sections: Section 1 asked for demographic details such as age, years of guiding experience and typical activities. The next section asked interviewees to detail their usual process for arranging activities with clients, and the usual format for summer and winter activities. The third section involved participants discussing their perception of hazards. The fourth section used the critical incident technique, asking participants to recount an adverse incident they had experienced, in as much detail as possible. The final section featured questions designed to further explore specific NTS (e.g. what are the key decisions you make during a trip?).

Analysis

Analysis of the interviews was conducted using theory-driven directed content analysis encompassing use of both deductive and inductive coding (Hsieh & Shannon, 2005). In practice this meant that a framework of generic NTS categories (situation awareness, decision-making, task management, leadership, teamwork and communication) was applied to the interview data, with a focus on content relevant to these skills. Inductive content analysis was also used to create original codes specific to the NTS behaviours within mountaineering context, and to produce NTS elements using a bottom-up approach. Factors influencing NTS were also coded using an inductive approach.

All coding was done using the Microsoft Word comment function by the first author, enabling text to be highlighted and then tagged with the representative code. Codes relevant to NTS were designed to concisely describe the behaviour outlined within the text and were attached to a broad skill category from the applied framework (e.g. 'checks weather' – situation awareness) where possible (some codes did not fit the framework and were used to form a new skill category). These codes were used to build a NTS framework specific to mountain guides, with all codes and elements checked and agreed by all authors. Codes relevant to influencing factors were simply designed to describe the relevant factor (e.g. 'fatigue').

Key results

The coding and categorisation process to date has produced six NTS categories with 18 associated elements, with work ongoing at the time of submission. This section will outline a selection of the key cognitive (Table 1) and then the social (Table 2) skills and elements identified to date. This will be followed by a brief overview of key influencing factors.

Cognitive NTS

The mountain guide interviewees highlighted the importance of gathering information prior to any activity, both in terms of the anticipated route or activity, and the status of their client. This was a necessity to ensure they had the information they needed for later decision-making related to client-route matching:

'Their technical skills are something that will help give them an enjoyable down the hill, but it also keeps them and yourself safe. So, making sure the route is appropriate to the client, I find this really important.' P4

This knowledge was compounded by both preparation activities designed to test client fitness and ensure all gear / kit was ready and appropriate for use, in addition to ongoing monitoring of the environment and client performance, to ensure actions were taken where needed to protect client safety and wellbeing:

'Every opportunity I have to look at them I look at them, just to see how they're moving, how they were moving before. If they're going slower, how they're breathing, how they're looking'. P1

The guides' knowledge of the mountains, including typical difficulty levels and timings for various routes, was utilised in their decisions at specific points during activities. Once a specific route was selected this led to a series of related decisions – with the chosen terrain dictating what kit might be needed, particular approaches etc. There were also various checkpoints along routes which guides used to facilitate their decisions on whether to continue with a certain activity and how long it might take:

'So there's a really classic sort of cut off time move, 2- 2.5 hours to this emergency hut on the Matterhorn. If that is really hard, if it takes you 2 hours 29 minutes, but you've been working really hard, it's going to be a very long day. If you get there in 2 hours 10 minutes it's all fine'. P2

Deciding to stop an activity was a critical, and often difficult, decision to make which was contingent on assessment of multiple factors including client fatigue, weather conditions and the perceived risk level.

Table 1: Selected mountain guide cognitive NTS categories, elements and codes.

Skill category	Element	Example codes
Situation awareness: *Building and maintaining an awareness of the environment, conditions and self. Recognising information in the environment, then using that to anticipate future states.*	**Gathering information:** *Seeking information relevant to the proposed activity / route / climb prior to activity.*	• Checks weather forecast. • Evaluates client physiological / mental condition.
	Maintaining / updating awareness: *Ongoing information gathering and collation during activity to consistently update mental model.*	• Regular visual / auditory checks of environment. • Monitors / checks on client movement / performance.
	Anticipating future events: *Predicting what might happen next according to both action and in-action.*	• Takes action to avoid adverse weather. • Anticipates mountain descent requirements.
Decision-making: *Reaching an appropriate judgement about a situation, selecting the most appropriate actions and managing risk.*	**Managing risk:** *Evaluating a situation to identify potential risks and hazards, to both self and clients, then acting to mitigate or remove those hazards.*	• Stops activity / action if conditions have increased risk (e.g. adverse weather, risk of avalanche). • Avoid specific points of terrain known to be hazardous / when conditions are hazardous.
	Identification and utilisation of decision-points: *Recognition of necessity of certain types of decision at specific points on a route / during an activity. Link between some decisions also acknowledged (if a then b).*	• Cascading decisions – route choice impacts equipment choice, impacts preparation requirements. • Staging decisions – decision to carry on with activity / route dependent on time taken to reach specified check point.
	Route selection: *Consideration of multiple factors, including client ability level, to select safest route for activity.*	• Engages in client - route matching – selecting routes / activity according to competence / training / tiredness.
Task management: *Organising activities and resources to maintain safety and quality standards. Managing competing pressures and demands.*	**Planning:** *Producing route and shelter requirements, locations and expectations prior to actual activity.*	• Manages timing of activities to enhance effectiveness / achieve objectives / avoid long waits on busy mountains. • Utilises knowledge of route / activity / trip time to plan day.
	Preparing: *Activities to ready self and clients for activities.*	• Checks and prepares client kit / equipment (including testing fit and utility). • Ensure clients have engaged in mental preparation for route.

Cognitive readiness: *Mental preparation and adaptability to cope with dynamic, mountaineering conditions.*	**Adaptation to conditions:** *Altering actions to suit changes in the environment, switch focus in response to environment and client conditions.*	• Continuously update plan to accommodate conditions and terrain. • Encourage / train clients to be flexible about plan and route.

Social NTS

Leadership was a core skill for mountain guides, with client groups reliant on guides for route finding and guidance on task activities alongside training and support when required. Guides highlighted that guiding clients through tasks could vary from the need for specific, technical direction, to overall monitoring of task and exertion levels to prevent client fatigue or discomfort:

'We try and slow people down generally. Like people will be over enthusiastic. And people even who've done – so if you do a four hour hut walk and you try and do it in three and a half hours, you'll be tired, whereas four and a half might feel easy'. P2

Despite the responsibility of mountain guides for decisions throughout activities, the majority of interviewees emphasised building an environment of psychological safety, where clients were provided with insights into the factors influencing decisions, were encouraged to share their opinions and engage in shared decision-making where possible:

'Create those pauses in your day, to have the discussions, and I suppose set a tone, where everybody's got a voice'. P9

This was linked to the importance of communication, and ensuring clients understood the reasoning behind key decisions and actions. This helped clients to feel involved in the various activities, and managed their expectations regarding upcoming exertion or required actions:

'You really need to explain quite clearly the summit I think, is that we're only halfway there. So this is going to take a lot of energy to get back down'. P2

Table 2: Selected mountain guide social NTS categories, elements and codes

Skill category	Element	Example codes
Leadership: *Supporting and managing client activities to achieve goals safely and emphasise engagement / involvement.*	**Directing / guiding activity:** *Guide client actions to complete the activity safely and effectively, while also engaging in training and education of clients where possible.*	• Manages activity exertion level to avoid over / under exertion. • Guide client actions directly to achieve activity goal.
	Developing psychological safety: *Ensure clients feel comfortable raising concerns / questions and communicating freely with guide.*	• Encourage group to make observations . ask questions. • Listen to concerns / worries / points raised by others.
Teamwork & Communication: *Sharing information, goals and understanding to facilitate group actions. Combining activities and effort to*	**Exchanging information:**. *Ensuring everyone is aware of the planned activity / route, is aware of the associated hazards / risks and knows the approach.*	• Pre-activity briefing (sharing weather, terrain, options, emergency protocols) with clients / team. • Discuss potential approaches / options prior to making decision.

reach a shared goal safely and effectively.	**Sharing and co-ordinating tasks:** *Interacting with others to ensure tasks are shared appropriately, and actions are co-ordinated to reach goal.*	• Co-ordinate client actions via visualisation (e.g. aim for that boulder). • Shares workload / tasks among team (e.g. shares task of carrying heavy load).

Influencing factors

Key factors influencing safety and performance of NTS included: fatigue, tunnel vision, competitiveness, task pressure and client behaviour.

Fatigue was commonly discussed with reference to over exertion, discomfort (physical fatigue), and the impact of fatigue on decision-making and attention (cognitive fatigue). All interviewees discussed the heightened risk associated with un-managed fatigue, with an associated rise in error rate:

'When I'm working if I'm tired, I'm careful with my decisions because I know I'm more prone or potentially prone to making weaker decisions'. P5

This was sometimes linked to tunnel vision, where attention became predominantly focused on a single element of a task, with exclusion of the broader context or environment. This focused attention was linked to reduced situation awareness and could result in members of a group going off-route, becoming left behind or not recognising changes within their environment:

'You might you be skiing with somebody who all their effort is going into the skiing and suddenly they look around and there's no group.' P2

External pressure could be conceptualised as competitiveness, between guides, groups and organisations, or task pressure linked to goal orientation – e.g. the need to reach a particular summit and satisfy client requirements. In both cases guides highlighted that this type of pressure could lead to risk taking, such as proceeding with an activity is less than ideal weather conditions, or rushing activities to 'beat' other groups:

'You tell yourself that you're making okay decisions about the snow and you're aware that you might not be because of other things like these clients are paying a lot of money and you'll want to go and do something'. P8

Finally, the clients themselves could also pose a risk to the group, and a barrier to performance of NTS. More specifically, clients were reported as disregarding instruction on occasion, being unprepared for particular activities, or might even be suffering from over-indulgence from the night before:

'I realized that my client was not that organised as he normally was. He has a habit of drinking too much alcohol, actually, and I think I was probably a little bit annoyed with him that I felt he could have been sharper in the morning'. P5

This could make it more difficult to anticipate client behaviour and as such could constrain decision making regarding certain approaches and routes.

Discussion

The results presented within the current paper contribute to the theoretical development of NTS by expanding the study of these skills into a new context, one with a recreational aspect. The identified elements and associated behaviours also broadening the range of elements encompassed within the generic framework of NTS and thus add to our understanding of these skills in practice.

The results emphasise the relevance and importance of NTS for mountain guides and provide the groundwork for development of a specific non-technical skills framework for guides. The results also identified factors influencing NTS, providing insight into the application of NTS within this dynamic and potentially dangerous context, as well as indicating a range of factors that may prove to be a barrier to effective NTS performance for mountain guides.

Analysis of the interview data aligned with the accepted generic framework of NTS categories (situation awareness, decision-making, task management, leadership, teamwork & communication) with the addition of cognitive readiness – a skill that is becoming progressively more recognised as an integral NTS (Irwin et al., 2023, Hamlet et al., 2021). The skill elements were developed using a bottom-up approach and as such are reflective of specific behaviours discussed by the mountain guide interviewees. Although some of these elements (such as gathering information and directing task behaviours) also feature in other NTS taxonomies (such as FLINTS – Irwin et al., 2022), the associated behaviours are unique to mountain guides. Other elements (such as identification and utilisation of decision points, and route selection) are entirely original to the current study. This aligns with guidance for developing behavioural marker systems for training and assessment of NTS, which should be tailored to each specific role and context to ensure the skill elements and behaviours are relevant (Yule et al., 2006). As such the current findings provide a useful baseline from which to develop a tailored NTS behavioural marker system for mountain guides.

The balance between professional expertise and client service was emphasised within the current results. Guides discussed monitoring both the environment and their client groups, then making decisions based on both aspects. This balance was made more difficult when influencing factors such as client behaviour and task pressure were in play, aligning with the risk factors identified by Girard and colleagues (2020). The pressure to achieve client satisfaction and a good experience can be linked to the concept of emotional labour, or ‘service with a smile’, where workers feel compelled to act in a way not necessarily reflective of their underlying emotions. Previous research with adventure guides reported that guides felt there were three key guiding responsibilities: safety, fun and building community. All three areas required emotion management, such as not showing fear within a risky situation, modelling enthusiasm for various activities and engaging in social facilitation (Sharpe, 2005). This emphasises the multiple pressures and expectations that guides may feel obliged to manage, and which may act as a stressor, with the potential to adversely impact their NTS during times of heightened pressure.

Fatigue management was an important aspect linked to multiple NTS within the current study, including the utilisation of situation awareness to evaluate, monitor and take action to remedy client fatigue levels, as well as directing activity to manage exertion levels as part of leadership. Most of the behaviours reported related to managing client fatigue, but many interviewees also highlighted the importance of self-management to ensure the guide remained at peak performance levels. This is a vital element of maintaining safety in the mountains, with participants within the current study highlighting the potential adverse effect of fatigue on both physiology and cognition. Fatigue has also been highlighted as a key causal factor within mountaineering accidents, primarily reported as fatigue caused by, or experienced within, the specific activity (Chamarro & Fernandez-Castro, 2009).

Limitations

The current study was based on self-report and recall of past adverse incidents., as such the results are subject to potential bias and errors related to reduced recall. The sample consists of British Mountain Guides, and as such the data does not necessarily generalise to other groups of mountain guides. Finally, ideally the data presented here should be validated via alternative methods of data collection (such as survey, observation or delphi methods) before being utilised within a behavioural marker system.

Conclusion

The current findings highlight the importance and utility of NTS within the context of mountain guiding, and identify six core skill categories (situation awareness, decision-making, leadership, teamwork & communication, task management, cognitive readiness) and associated elements. This suggests a need for the development of tailored NTS training for guides to support and enhance safety, ideally in conjunction with guidance regarding factors, such as fatigue and task pressure, that may influence the performance of these skills in practice.

References

Catchpole, K., Mishra, A., Handa, A., & McCulloch, P. (2008). Teamwork and error in the operating room: analysis of skills and roles. *Annals of Surgery*, *247*, 699-706.

Chamarro, A., & Fernández-Castro, J. (2009). The perception of causes of accidents in mountain sports: a study based on the experiences of victims. *Accident Analysis & Prevention*, *41*(1), 197-201.

Fjeld, G. P., Tvedt, S. D., & Oltedal, H. (2018). Bridge officers' non-technical skills: a literature review. *WMU Journal of Maritime Affairs*, *17*, 475-495.

Flin, R., O'Connor, P., & Crichton, M. (2017). *Safety at the sharp end: a guide to non-technical skills*. CRC Press.

Girard, A., Caroly, S., & Falzon, P. (2020). Mountain guides' everyday work: Articulating safety and service relationship. In. *Contemporary Ergonomics and Human Factors 2020,* Eds: R. Charles & D. Golightly.

Hamlet, O. E. D., Irwin, A., & McGregor, M. (2020). Is it all about the mission? Comparing non-technical skills across offshore transport and search and rescue helicopter pilots. *The International Journal of Aerospace Psychology*, *30*, 215-235.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277-1288.

Irwin, A., Tone, I-R. & Sedlar, N. (2022). Developing a Prototype Behavioural Marker System for Farmer Non-Technical Skills (FLINTS), *Journal of Agromedicine*, DOI: 10.1080/1059924X.2022.2089420

Naweed, A., & Murphy, P. (2022). One-track mind: investigating positive and negative applications of non-technical skills in rail network control. *Applied Ergonomics*, *106*, 103840.

Reader, T., Flin, R., Lauche, K., & Cuthbertson, B. H. (2006). Non-technical skills in the intensive care unit. *BJA: British Journal of Anaesthesia*, *96*, 551-559.

Rebelo, C. F. C., Ezequiel, G. M. G., Mendes, S. L. D. C. M., & Carvalho, M. J. P. D. J. (2018). It is All about Safety: An Experience in Pico Mountain—Portugal. *Tourism Planning & Development*, *15*, 134-148.

Sharpe, E.K. (2005). 'Going above and beyond'. The emotional labour of adventure guides. *Jounral of Leisure Research, 37,* 29-50.

Sneddon, A., Mearns, K., & Flin, R. (2006). Situation awareness and safety in offshore drill crews. *Cognition, Technology & Work*, *8*, 255-267.

Stewart-Patterson, I. (2004). Decision making in the mountain environment. In *International Snow Science Workshop Proceedings*, 535-545.

Walker, E., & Latosuo, E. (2016). Gendered decision-making practices in Alaska's dynamic mountain environments? A study of professional mountain guides. *Journal of Outdoor Recreation and Tourism*, *13*, 18-22.

Using SUS for Current and Future AI

Richard Farry[1]

[1] QinetiQ

SUMMARY

The System Usability Scale (SUS) was assessed for its relevance and ease of use for assessing an AI capable of human-like interaction. Participants used SUS to assess Outlook, a contemporary consumer-grade AI interaction partners (smartphone digital assistants), and human teammates as a proxy 'system' for future human-like AI interaction partners. The results show that participants considered SUS to be relevant and easy to use for contemporary consumer-grade AI interaction partners, but not for human teammates. However, there was no meaningful difference in their ability to apply SUS between contemporary digital assistants, human teammates, and an email client. Thus, SUS can be used effectively for all of these kinds of systems.

KEYWORDS

SUS, Usability, AI, Human-Autonomy Teams

Introduction

Current and future Artificial Intelligence (AI) systems, particularly those intended to interact with humans as part of a human-autonomy team, will need to be assessed for their usability. It is not known whether current usability assessment methods are or will be suitable to assess such systems, particularly as (or if) AIs become more human-like in their interaction roles and competence.

This study set out to investigate the suitability of the System Usability Scale (Brooke 1996) to evaluate contemporary AI interaction partners, and future AI systems. A key strength of SUS, and why it was selected for this study, is that it can be used to assess a broad range of systems from any domain (Stanton *et al.* 2005), due to its use of general and high-level statements for participants to respond to. Additionally, SUS is easy to administer, can be used with small sample sizes with reliable results, and is valid; able to discriminate between usable and unusable systems (Brooke 2013).

Method

The hypothesis to be tested was whether participants found the SUS as relevant and easy to use for current and future AI systems that are intended to interact in a 'natural' way with humans, compared to using SUS for 'classic' desktop software using a Windows Icon Mouse Pointer (WIMP) interface. As future AI systems with human-like interaction were not available at the time of carrying out the study, human teammates were used as a proxy for such systems. The study participants completed a questionnaire that included three rounds of SUS. In order, they were for 'classic software' (Microsoft Outlook; the email client used within the participants' organisation), current and commonly available AI-based Digital Assistants (Alexa, Google Assistant, or Siri), and future highly-capable and human-interaction-like AI systems (using Human Teammates as a proxy for these future systems).

Following completion of each round of SUS the participants were asked to respond to the following statements on a 5-point Likert scale (from ‘Strongly Disagree’ to ‘Strongly Agree’): “The [SUS] statements were relevant to the system”, and; “I found it easy to rate the system”. The participants were then invited to provide feedback about their experience of using SUS for each system.

Results

The participants were asked for each of the systems whether they considered the SUS statements to be relevant. The results are shown in Figure 1 below. There were thirty-one participants in total, but only eighteen of the participants responded to the questions about Digital Assistants.

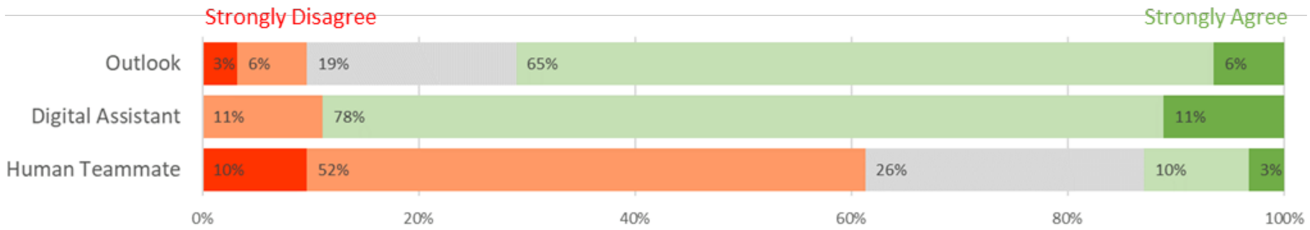


Figure 1: "Statements were relevant to the system" (n = 31, 18, 31)

The results indicate broad agreement that the SUS statements were relevant in the case of Outlook (71%, n = 31) and Digital Assistants (89%, n = 18). However, the participants considered the statements less relevant for their human teammates (agreement 13% and disagreement 62%, n = 31).

The participants were asked for each of the systems whether they found it easy to rate the SUS statements. The results are shown in Figure 2 below.

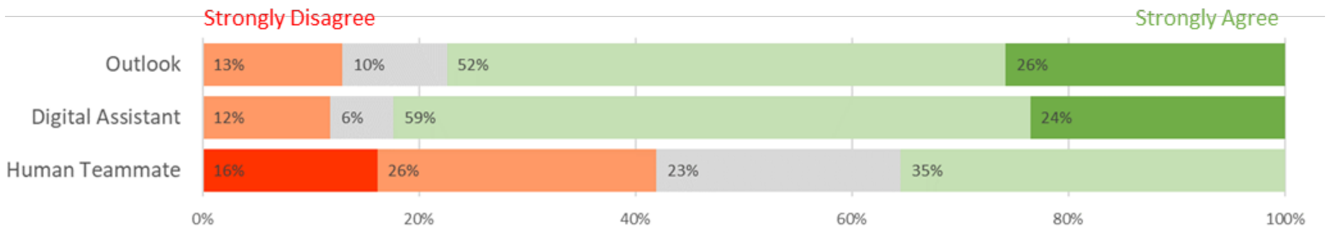


Figure 2: "I found it easy to rate the statements" (n = 31, 17, 31)

The results indicate broad agreement that it was easy to rate the SUS statements for Outlook (77%, n = 31) and Digital Assistants (82%, n = 18)[1]. However, there was less agreement and more disagreement on whether it was easy to rate the SUS statements when it came to their human teammates (agreement 35% and disagreement 42%, n = 31).

To investigate the objective usage of SUS a comparison of the proportions of ‘Neither Agree or Disagree’ ratings was carried out between Outlook and Digital Assistants, and between Outlook and Human Teammates. For this purpose an equivalence test (Lakens et al. 2018) was carried out, and in both cases the proportion of ‘Neither Agree or Disagree’ ratings were found to be equivalent

[1] The minor discrepancies between the percentages in the text and the figure for total agreement (agree plus strongly agree) are due to rounding.

(Digital Assistant: effect size tested = 0.1, $Z = 2.483$, $p < 0.01$. Human teammate: effect size tested = 0.1, $Z = -2.147$, $p < 0.05$).

SUS Scores

The overall SUS scores for each system are provided in Figure 3 below.

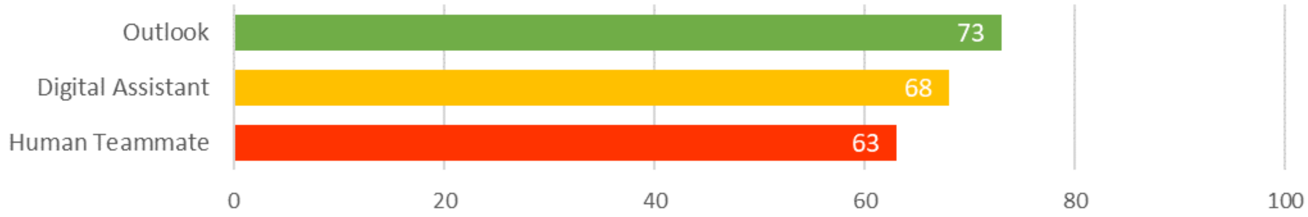


Figure 3: SUS Scores (68 is considered an average usability score (Lewis 2018))

Comments

The participants were asked to provide feedback on the use of SUS for each of the three systems. Twenty-four comments were received, and of these fifteen were about the use of SUS. The remaining comments related to the system being assessed. The comments about using SUS are summarised in the table below.

Table 1: Summary of comments about the use of SUS

System	Comment Type/Category
Outlook	SUS is too generic to capture relevant usability feedback (n = 3)
Outlook	Outlook has so much functionality [of varying usability] making it difficult to know how to respond to the questions (n =2)
Digital Assistant	It was more difficult to use SUS for a voice-based interface than a 'point and click' interface [i.e. a Windows Icons Mouse Pointer (WIMP) based interface] (n = 1)
Digital Assistant	The Digital Assistant is a front end to a range of functions / other systems, so it was unclear how to respond (n = 1)
Human Teammate	The SUS questions were difficult or not relevant to humans (n = 3)
Human Teammate	The SUS questions were impossible to answer about humans (n = 1)
Human Teammate	Neutral comment about the appropriateness of using SUS for a human (n = 1)
Human Teammate	Positive comment about the appropriateness of using SUS for a human (it was described as 'hilarious') (n = 1)
Human Teammate	Negative comment about the appropriateness of using SUS for a human (it was described as 'not appropriate' and 'demeaning') (n = 2)

Note that the participants who said rating the human teammate was difficult or impossible to do all successfully completed SUS, though their selection of the 'Neither Agree or Disagree' ratings (which could indicate difficulty in responding or simply just giving a neutral rating) accounted for 30.6% of their answers. The overall rate of participants responding with 'Neither Agree or Disagree' was 17.5% for Outlook, 15.3% for Digital Assistants, and 21.3% for Human Teammates.

Discussion

The participants rated Outlook as more usable than Digital Assistants and Human Teammates. This is perhaps not surprising in that Outlook is a tool designed to be usable, and is understandable and predictable in terms of its design and intended function (using what the philosopher Daniel Dennett refers to the as the 'design stance' (Dennett (2009)), whereas teammates are not, are far more complex, and unlike tools have other interests and goals.

Overall the participants considered SUS to be a relevant and easy to use tool to assess contemporary AI interaction partners (in the form of Digital Assistants), but not human teammates (as a proxy for future AI interaction partners). However, it was found that their ability to respond positively or negatively to the SUS statements for both was equivalent to using SUS for 'classic' software (in this case, Outlook). Thus, while subjectively they did not consider SUS to be valid, in practice their use of SUS demonstrated that it is an effective tool to assess AI systems, including those capable of human-like interaction.

Of more concern are some of the negative comments received about referring to people as systems, including one participant considering it to be 'demeaning' (see Table 1). It is unclear at this time whether similar concerns will arise for future AI systems, but it seems likely that they will if such systems are sufficiently anthropomorphic (or zoomorphic) or promote emotional engagement or attachment. This negative aspect of the use of SUS might be ameliorated with an appropriate briefing or introduction

SUS should be considered an appropriate means to measure the usability of contemporary AI interaction partners, and an effective stop-gap for measuring the usability of more advanced AI systems.

References

Brooke, J. (1996) 'SUS: A "quick and dirty" usability scale', in Jordan, P., Thomas, B. and Weerdmeester, B. (Eds.), *Usability Evaluation in Industry*, pp.189-194. Taylor & Francis.

Brooke, J. (2013) 'SUS: A Retrospective", *Journal of Usability Studies*, Vol. 8, No.2, pp.29-40.

Dennett, D. (2009) 'Intentional Systems Theory'. The Oxford Handbook of Mind, p339-350.Lakens, D., Scheel, A., and Isager, P. (2018) 'Equivalence Testing for Psychological Research: A Tutorial', *Advances in Methods and Practices in Psychological Science*, Vol.1, No.2, pp.259-269.

Lewis, J. (2018) 'The System Usability Scale: Past, Present, and Future', *International Journal of Human-Computer Interaction*, Vol.34, No.7, pp.577-590.

Stanton, N., Salmon, P., Walker, G., Baber, C. and Jenkins, D. (2005) *Human Factors Methods: A Practical Guide for Engineering and Design*. Ashgate.

Harnessing A Human Factors Approach to Improve Patient Safety

Jenny Sutcliffe[1], Suzi Lomax[1] & Jennifer Macallan[1]

[1]Royal Surrey NHS Foundation Trust, UK

SUMMARY (for short papers, 2 pages max)

The interest in employing Human Factors (HF) in healthcare is increasing. The SCReaM HF and Team Resource Management (TRM) programme is aimed at raising the awareness, understanding and application of the science of HF within healthcare to help staff improve their safety and wellbeing and that of their patients. The programme is divided into three strands: rolling training, HF Projects and HF Engineering. The programme has been successfully embedded into an NHS Trust and provides a good model for how HF can be introduced and utilised within healthcare.

KEYWORDS

Healthcare, Patient Safety, Systems Thinking

Introduction

SCReaM Human Factors (HF) and Team Resource Management (TRM) is a programme developed to improve patient/staff safety and wellbeing through the understanding and application of HF. HF in healthcare is often misunderstood, having been historically based on crew resource management (CRM) training adapted from aviation.

However the true scope of HF in healthcare reaches far beyond this and lies in understanding and applying systems thinking methodology in this complex adaptive industry. The programme encapsulates CRM principles, Quality Improvement (QI) methodology, and HF methodology to enable staff to design their system to best fit their ways of working and improve theirs and their patients' safety and wellbeing.

The programme is divided into the strands: A rolling training programme; HF projects, primarily stemmed from recurring 'pledge' themes; and HF Engineering, the provision of our HF expertise to Trust-wide Transformation programmes.

Background

The SCReaM programme (formerly Surrey Crisis Resource Management) began in 2013 when two anaesthetists identified that during emergencies key lifesaving steps are omitted because human memory and performance are negatively affected by stress. They developed and introduced the first set of UK emergency prompt cards for operating theatres. These cards were supported with the introduction of multidisciplinary training available to all theatre healthcare workers. In December 2018, the word 'Crisis' was removed from the already established brand and the programme re-focussed on the proactive, rather than reactive, nature of HF. A rolling training programme was created to ensure staff could maintain up-to-date knowledge and skills, and 'pledges' were introduced into the courses. Pledges, based on the PDSA (Plan-Do-Study-Act) cycle, are small simple changes delegates make to their work system based on something they learnt on the course.

Training Programme

The training is accredited and advertised by the Royal College of Nursing (RCN), Royal College of Surgeons (RCS), Royal College of Anaesthetists (RCoA) and the Chartered Institute for Ergonomics and Human Factors (CIEHF), and the Clinical Human Factors Group (CHFG). This, along with the option to attend courses virtually or in the classroom, means delegates from other Trusts can easily attend the courses; for which there is a continual uptick in demand.

To-date over 107 courses have been carried out, filling over 1117 training spaces. Courses have been run in the ED, cardiology, aseptics, and radiology departments, as well as the newer rolling programmes running in theatres and oncology. Courses are multidisciplinary, tailored to each department, and are facilitative. They cover CRM principles as well as specific HF concepts such as the SHEEP model (Rosenorn-Lanng, 2014), varieties of human work (Shorrock, 2016), three-models of safety (Vincent and Amalberti, 2016), and the importance of systems thinking.

A key indicator of success is that course delegates understand what HF is and its impact. Post-course questionnaires suggest delegates are accurately taking away the key messages, with 80% including key phrases when asked to describe what HF is, models of safety, teaming, and stress.

HF Projects

To-date 515 pledge ideas, resulting from the training, have been created, with 223 (42%) of these successfully completed leading to improved safety and wellbeing. Some pledges related to enhanced quality of care: improving the team brief, identifying antibiotic requirements, and reducing multi-tasking. Other pledges addressed patient experience of care: better tracking of allergies, introducing the team to patients before treatment, minimising distractions during anaesthetising. Others related to the staffs' ability to manage that care: notably, since Covid-19 hit, pledge themes dramatically skewed towards Stress & Wellbeing (from ~4% pre pandemic to ~25% now), specifically focussing on breaks, providing support groups, and actively living by the values of kindness and civility. Of the completed pledges, 76% stated they would continue to "do it like this from now on" or "try it again… twice". This demonstrates that the delegates are finding success in their improvements.

HF Engineering

Having a Chartered Ergonomist integrated into the Trust's Transformation team has meant the Trust is starting to better understand the symbiotic nature of HF and QI. Notably the Trust has started to utilise HF to support the roll out of the new Patient Safety Incident Response Framework (PSIRF) as part of the patient safety incident response system. Elsewhere in the Trust the SCReaM team have been used to improve staff escalation in maternity; creating the TEACUP framework for escalation and identifying six people & system-based interventions.

Conclusion

This programme is unique: there is currently no 'best practice' guidance for integrating HF into healthcare. No other NHS Trust has a rolling HF training programme as part of their business as usual strategy, nor have any employed a Chartered Ergonomist integrated into the Trust's Transformation team to support wider transformation programmes. The benefits of utilising HF are well documented, but when it comes to providing quantitative evidence that budget holders often need to justify expenditure, evidence can be thin on the ground. This programme will continue to build on the good work already undertaken, record its successes, and share these amongst the HF community to help others integrate HF into healthcare.

References

Rosenorn-Lanng, D. (2014). Human Factors in Healthcare: Level One. Oxford: OUP.
Shorrock, S. & Williams, C. (2016). Human Factors and Ergonomics in Practice. CRC Press.
Vincent, C. & Amalberti, R. (2016) Safer Healthcare: Strategies for the Real World. NY: SpringerOpen.

Consideration of Stakeholders for Technology Acceptance in Marine Conservation

Dr Ella-Mae Hubbard[1], Melissa Schiele [2] & Prof. Paul Lepper[3]

[1]Human Factors Engineering Research Group, Intelligent Automation Centre, Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, [2]Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University and Institute of Zoology, ZSL, [3]Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough

SUMMARY

This paper discusses the importance of understanding various stakeholder perspectives before investigating technology acceptance. Stakeholders are identified, from a systems perspective, with their key requirements and interactions. This is part of wider work towards developing an Augmented Technology Acceptance Model (for example acceptance of drone technology) within the context of marine conservation.

KEYWORDS

Technology acceptance, stakeholders, marine

Introduction

Digital technology is developing at speed and application areas are constantly emerging. For such technology to be successfully implemented, we need to appreciate the perception and potential impact for all stakeholders (not just operator/ user/ owner/ customer). Without acceptance from all stakeholders, operational efficiency is incidental. Technology acceptance is not a new field of research (Davis, 1989; Venkatesh and Bala, 2008), however, it is clear that some sectors have a more mature understanding. For example, whilst there is some appreciation of impacts of technology within conservation (Hahn et al, 2022; Speaker et al, 2021; Wilfred et al, 2019), the marine sector has not considered acceptance of digital technologies in any formal depth. The overall aim of this work is to understand barriers and drivers which may oppose or support technology acceptance of digital technology (e.g. drones) within the marine conservation context, applications such as wildlife surveillance, ocean contamination detection, and legal and illegal operations at sea.

Stakeholders

The importance of technology acceptance in this context became evident during a variety of field work studies, which were initially focused on design and deployment of digital technologies within marine conservation. During these field work studies, stakeholders were identified and mapped (using stakeholder influence techniques) using an ethnographic approach.

This enabled the understanding of the variety of stakeholders who have an interest or impact in digital technology in a marine context. A summary of this is presented in Table 1.

Table 1: Stakeholders views

Stakeholders	System involvement	Key concerns
Manufacturer	Produces digital technology	Will people know how to use the system?
Operator	Operates digital technology	How do I operate the system? What rules should I be aware of?
Local industry and Charitable organisations	May be able to implement the technology	Need to understand how we can use/ exploit the technology. What are the opportunities and risks?
Security (private and public)	Maintain security and safety	Are these systems being used appropriately? What is the potential for misuse?
Regulatory bodies	Provide/ maintain relevant regulations and guidance (for the technology and operators)	Do our current regulations cover any new technologies?
Local residents	Live and/ or work in areas where the technology may be implemented	Will this impact my quality of life? I'm worried because I don't really know how it works.
Students and educational establishments	Learning about the technology being implemented and the contexts in which it is deployed.	Exciting, applied opportunities to learn. How do we keep the system up to date?
Research scientists/ conservation technologists and engineers	Utilising the technology to gather data in fieldwork	How much can I trust the system and the data it generates? Does the technology help or make my life harder? Does this change the job I do? Will my team make use of the technology or is it just a waste of money?
Maintenance team	Maintaining the technology	Will I get the training and support I need if this is not technology I am familiar with?
Visitors	Visiting the areas where the technology may be implemented	Will this positively or negatively affect my visit? Will it cost me money?

Conclusion – next steps

Identification of relevant stakeholders is a necessary first step towards developing an approach to technology acceptance. It is important to consider acceptance by all stakeholders within the system (or ecosystem). Barriers and drivers for technology acceptance should be considered as early as possible in the process – not just when preparing for implementation. Next steps for this work is to engage with representatives from the identified stakeholders (through observation, questionnaires and semi structured interviews) to develop the Augmented Technology Acceptance model. Further stakeholder interviews and field work will also be used to validate the model. The understanding gained from this stakeholder analysis has allowed for targeted design of this data collection and validation. Parallel work is considering transferring lessons learned across industrial sectors (for example, marine conservation, manufacturing and defence). The output from this work will help enhance the business case and inform management practices to support successful adoption of digital technology. It will also go on to understand the potential impacts on all stakeholders, supporting a human centred view on design optimisation and suitable deployment of digital technology.

References

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 319–340. https://doi.org/10.2307/249008

Hahn, N.R., Bombaci, S.P. & Wittemyer, G. Identifying conservation technology needs, barriers, and opportunities. *Sci Rep* **12**, 4802 (2022). https://doi.org/10.1038/s41598-022-08330-w

Speaker, T., O'Donnell, S., Wittemyer, G., Bruyere, B., Loucks, C., Dancer, A., Carter, M., Fegraus, E., Palmer, J., Warren, E., & Solomon, J. (2022). A global community-sourced assessment of the state of conservation technology. *Conservation Biology*, 36, e13871. https://doi.org/10.1111/cobi.13871

Venkatesh, V. and Bala, H. (2008), Technology Acceptance Model 3 and a Research Agenda on Interventions. Decision Sciences, 39: 273-315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

Wilfred, P, Kayeye, H, Magige, FJ, Kisingo, A, Nahonyo, CL. Challenges facing the introduction of SMART patrols in a game reserve, western Tanzania. *Afr J Ecol*. 2019; 57: 523– 530. https://doi.org/10.1111/aje.12634

Human-robot interaction: Assessing the ergonomics of tool handover

George V. Papadopoulos[1], Michail Maniadakis[1]

[1]Foundation for Research and Technology Hellas, Heraklion, Crete, Greece

SUMMARY

This work focuses on human-robot collaboration for assembly tasks, examining the position of robot-to-human handover of objects. A simulation environment is implemented to ergonomically evaluate the expected posture of the human arm in hypothetical delivery positions in the 3D space.

KEYWORDS

Human-Robot Interaction, Tool Handover, Arm Posture

Introduction

Collaborative robots (cobots) are becoming popular in automotive assembly industry because they can help improve productivity of assembly workers (Ajoudani et al, 2018) [1]. Employees performing physically demanding or repetitive tasks, can be assisted by cobots to improve their working conditions and reduce the risk of musculoskeletal disorders.

A common task that cobots may undertake in collaborative assembly applications is bringing tools from locations away from the assembly location and handover them to the workers. Particularly in assembly environments, tool-passing can happen hundreds of times during the day, and the relevant workload may strain arm muscles. Accordingly, the choice of tool handover position is an important decision made by the robot because, if done with ergonomic criteria, it can reduce the negative impact on the worker's muscles and in the long run it may significantly reduce the occurrence of musculoskeletal problems.

The current work focuses particularly on the study of tool handover, proposing a method that enables the robot to proactively examine the expected posture of the human arm and choose handover positions that “drive” the human arm to ergonomically suitable postures (Vianello et al, 2021). To this end, a simulation environment is implemented where several candidate delivery positions can be evaluated against known ergonomic criteria, considering the expected human arm movements. The robot then selects handover points that enforce ergonomics for the human arm.

Data Collection and Research Design

Demographical data, including age, gender, height, weight, and dominant hand, was collected from a sample of assembly workers. These data were used to adjust the human model in the simulation environment to match the somatometric characteristics of each participant. The study was designed as a simulation-based evaluation of human arm postures during tool handover.

Simulation Environment

The simulation of human actions is implemented in the PyBullet environment (Coumans). We use a human model with 25 degrees of freedom, which can be adjusted to match the somatometric characteristics of any given worker. Moreover, by using the profile of the worker the robot knows whether he is left- or right-handed and examines tool-handover at the appropriate arm. The simulation environment facilitates the use of constraints on arm motion, to simulate realistic hypothetical movements of the human arm, which are then evaluated against ergonomic criteria.

Ergonomic Assessment

The ergonomic criteria provided by the German Institute for Occupational Safety and Health (DGUV) are used to assess and categorize arm postures in three levels: (i) ergonomically appropriate, (ii) conditionally acceptable, and (iii) ergonomically unacceptable, as summarized in Table 1. We examine the angles of four degrees of freedom for the expected human upper limb posture, three at the shoulder and one at the elbow. Then overall ergonomic score of the given arm posture, equals to the score of the angle assessed with the lowest ergonomic score.

When the robot interacts with a human we randomly generate candidate positions/points and examine the hand poses (Figure 1). After finding 10 positions that meet all ergonomic criteria, we select the one that is closer to the robot, and this is used to implement the handover of the object.

Table 1: Ergonomic criteria for the arm

	Ergonomically appropriate	Conditionally accepted	Unacceptable
Shoulder abduction/adduction (ang0)	-20° <ang0< 0°	-60°<ang0<-20°	ang0<-60° or 0°<ang0
Shoulder flexion/extension (ang1)	0° <ang1< 20°	20° <ang1< 60°	ang1<0° or 60°<ang1
Shoulder rotation (ang2)	-15° <ang2 <30°	-30°<ang2<-15° or 30° <ang2 <60°	ang2 <-30° or 60° <ang2
Elbow flexion/extension (ang3)	60° <ang3 <100°	-	ang3 <60° or 100° <ang3

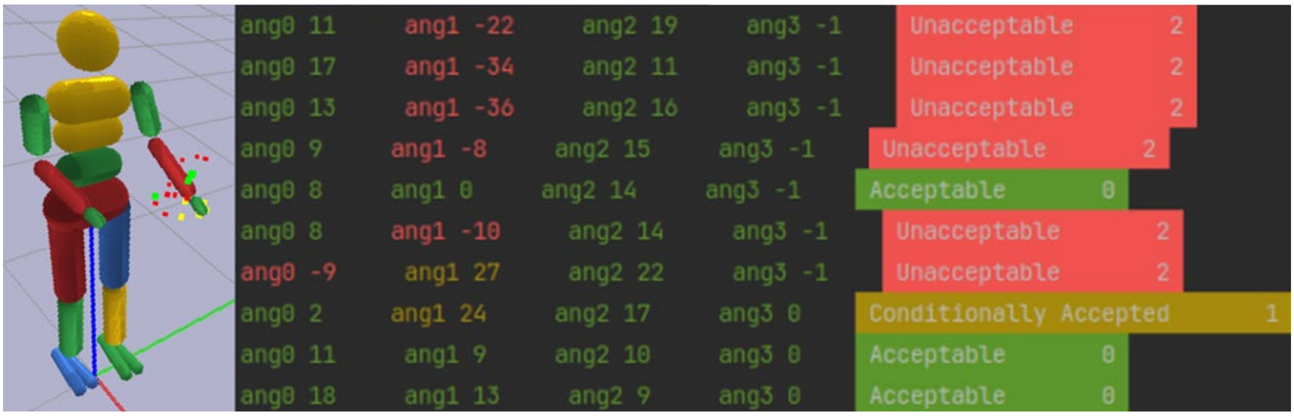


Figure 1: The ergonomic assessment of randomly generated candidate handover positions.

Discussion and Implications of Results

The simulation results show that the proposed method can effectively evaluate the expected posture of the human arm during tool handover and choose a delivery position that meets ergonomic criteria. This can reduce the negative impact of daily work activities on the worker's muscles and reduce the occurrence of musculoskeletal problems in the long run. Future work in this area could incorporate safety measures and evaluate the proposed framework in the real world.

Acknowledgment

This work has received funding from the European Union’s Horizon 2020 Research and Innovation program under grant agreement No 101017151, FELICE-Project.

References

Ajoudani A, Zanchettin AM, Ivaldi S, Albu-Schäffer A, Kosuge K, Khatib O (2018) Progress and prospects of the human–robot collaboration. Auton Robots 42:957–975

Vianello L, Mouret J-B, Dalin E, Aubry A, Ivaldi S (2021) Human Posture Prediction During Physical Human-Robot Interaction. IEEE Robot Autom Lett 6:6046–6053.

E. Coumans and Y. Bai, “Pybullet, a python module for physics simulation for games, robotics and machine learning,” https://github.com/bulletphysics/

DGUV Information 215-210 „Natürliche und künstliche Beleuchtung von Arbeitsstätten“. 64

Human performance in the rail freight yard

David Golightly[1] James Lonergan[2] David Ethell[3]

[1] School of Engineering, Newcastle University, [2] Rail safety and Standards Board, [3] National Freight Safety Group

SUMMARY

Human performance in the rail freight yard has been identified as source of risk to rail freight operations. It is, however, an area that has received little research attention. Observations and expert elicitation explored work in the freight yard, leading to an understanding of freight yard activities, the impact of freight yard design and environment, and external pressures. Together, these factors make the freight yard a complex and challenging environment, where fluid, cognitive planning optimises physically demanding, cooperative processes. The implications for future management of rail freight operations are discussed.

KEYWORDS

Rail human factors; freight; logistics; planning

Introduction

Rail freight is a key function of the economy. Freight moves bulk goods such as aggregates and fuel, intermodal containerised goods, dangerous goods such as nuclear fuel, and providing supplies and train movements for the build and repair of the railways itself. In Great Britain (GB), the total economic and social benefits of freight are valued at £2.5bn annually and removes the equivalent of 7 million heavy good vehicles from the roads (Rail Delivery Group, 2021). Therefore, the continued success and growth of rail freight is a cornerstone of transport decarbonisation, nationally and globally (e.g. UNESCAP, 2021).

Most importantly, the carriage of freight needs to be safe, ensuring the integrity of the load, and safety and staff and public. Rail freight also needs to be reliable. Incident-free rail freight is essential to ensure existing freight customer confidence while attracting new customers. Delays to freight trains can be costly, with minor incidents costing thousands of pounds in delay costs, through to accidents that might involve the loss of the freight load, damage to infrastructure or potentially weeks of disruption to both passenger and freight services (e.g. RAIB 2022).

In Great Britain, the 2020 Rail Safety and Standards Board (RSSB) Annual Health and Safety Report highlighted that in the previous two years, there had been a rise in the number of potentially higher risk train accidents for freight, a trend driven by an increase in derailments. Further, over this period, 288 trains were stopped on the network due to issues with vehicles, importing safety risks and delays to the network. The National Freight Safety Group (NFSG) has been set up to address rail freight risks. NFSG has identified that the condition of vehicles entering the network is the highest priority risk for the freight community, and is currently sponsoring a project to understand why freight vehicles may enter the rail network in an unsafe condition. The RSSB and Newcastle University are supporting the work undertaken in this project. This involves developing a better understanding of the processes prior to a vehicle entering the network and the underlying causes that may be a precursor to vehicles entering unsafely.

In a structured analysis of rail freight incidents on the network (Golightly et al., 2022), a number of human performance issues were identified that caused or contributed to events linked to the condition of the freight vehicle on the network, or in the freight yard. Typical events involved runaways in or outside of the yard, handbrakes left on wagons or airbrakes left on locomotives which then damage the wagon and rails if the train departs onto the network, or wagons entering the network in an unfit state (e.g. poorly loaded leading to derailment, damaged parts hanging out of gauge leading to collision). Slips, lapses and omissions in train preparation were the major human performance issues identified in yard tasks. The analysis also explored potential causal or contributory factors that led to these events. Usability of equipment, yard conditions (lighting, walking routes), wagon maintenance condition, time pressure and the organisation of work were identified as underpinning factors behind these human performance issues.

These initial findings warranted further exploration both to further understand the causal mechanisms, and to identify solutions. However, from a human factors perspective, freight functions such as the management of wagons in yards are one of the most under-researched areas of rail operations (Ryan et al., 2021). Human factors knowledge of tasks, competences, immediate and wider work environment, and pressures due to cultural / commercial / policy constraints is not widely available. Zhang et al (2019) carried out a structured analysis of US freight train accidents, and identified the preeminent types of accident were derailment and collision, with a range of human performance factors as primary causes. However, these human factors causes are attributed as a direct cause of the accident as it occurred (primarily attributed to the driving role) and the analysis did not look back into causes or human performance failures that may occur in the yard potentially leading to issues out on the network. Lawton (1998) studied violations in shunting work in freight yards. While the focus was somewhat different from Golightly et al (2022) (e.g. the work was pre-privatisation and therefore the organisation of work was different; Golightly et al (2022) found few violations in their dataset), many of the factors influencing work (e.g. time pressure, work arounds to complete tasks quickly) were similar. While Bowler and Basicik (2015) study rail operations at a port, their outputs are methodological, though they do state the difficulty in understanding work purely from procedures and the importance of observations and interviews to understand risk. Vaghi et al (2018) do not present human factors findings *per se*, but do highlight the importance of understanding human factors as an influencing factor in the deployment of new technology for rail freight. Hricova (2016) gives an example of this, highlighting the benefits of RFID tags on wagons to reduce error in wagon identification.

Given the significant knowledge gap in how work is performed in the freight yard, the following study aimed to develop knowledge of freight yard practices, in order to evolve our understanding of how human factors in freight yard work may contribute to freight train incidents on the network. Specific objectives included (1) capture freight yard tasks and activities; capture the environmental and design aspects of the freight yard (2) identify specific human performance risks (3) identify future steps to address human performance risks. Critically, and following on from the understanding of people as contributors to safety, as much as potential points of failure (Ryan et al., 2022) we also sought to understand the factors that might facilitate work, and how people adapted their work to fit conditions and constraints.

Method

The method involved a two-stage approach, building on the groundwork and understanding of the freight yard context in Golightly et al (2022). The first stage involved site visits with both observations, and structured and unstructured interviews with operational and management staff. This involved visits to five different freight yards and constituted over 35 hours of observations. During the course of these observations, informal interviews took place with over 30 members of staff on site, across multiple grades and functions. Observations included supervisory areas in freight yards for operational planning, extensive walk-arounds of the yard involving observations of freight preparation activities, observations of maintenance work, train cab access and opportunities to try out train preparation tasks. This included one observation during the night shift (a time of high workload in that particular yard) to observe conditions at that time of day. Observations also included visits to office areas for freight commercial planning, as this gave important insight into the inputs that shaped work in the freight environment.

Contemporaneous notes were taken, with a debrief between the authors after each visit. The observations lead to summary materials including a site complexity risk mapping, task models, and presentation. These materials were then used in a validation workshop with 13 members of the rail freight community. This workshop lasted a full day and involved structured activities to (1) validate task models (2) confirm key human factors challenges (3) explore strategic solutions. Comments were collated in breakout groups through the use of 'workbooks' where participants were guided through questions or to complete tables relevant to the discussion of freight yard activity. Notes arising from each theme were analysed and summary conclusions were drawn.

All work was conducted under Newcastle University Ethics 22-030-GOL.

Results

Work in the Yard

Typical yard tasks include receiving and stabling trains, moving wagons and locos, composing new train sets, preparing wagons for the network (e.g. preparing couplings, checking handbrakes), inspecting wagons, and negotiating with either the mainline rail network, or the yard of a receiving customer (e.g. a port) to dispatch a departing train. Key roles were the drivers (either mainline drivers or shunters), supervisors (who planned day to day operations), groundstaff (who prepared trains and wagons, amongst a range of other tasks), and maintenance staff. The configuration of roles changed due to local practice, needs and resourcing. One yard involved remote supervision; in another the supervisor conducted groundstaff work. Yards visited noted different peaks of work depending on the type of freight handled (e.g. night shifts at a yard linked to a steel works; approaching weekends when preparing infrastructure engineering trains).

In terms of the yard itself, we found an environment that was complex, physically and organisationally. Capacity was often limited as 1) specific tracks (or roads) in the yard had designated purposes (e.g. for refuelling) preventing flexible use 2) wagons were stored for maintenance 3) certain trains being prepared or stabled needing to be split to fit within yard constraints. Therefore, yards that seemed large were often very restricted in capacity. Track length is also an issue as this required longer trains to be split for stabling overnight, and then rejoined when being prepared for departure.

The number of movements coming into and out of the yard could be high, with trains arriving every few minutes. These might be trains specifically for processing in that yard, but also when locomotives were needing to find somewhere to stable during mechanical failures, or during rest breaks for drivers. Furthermore, the movement of trains *within* yards was often high, to construct train consists, move wagons for maintenance and so on. This increased the physical risk associated with moves as well as the number of times handbrakes needed to be applied or released – a key problem when trains with handbrakes missed went out on the network. In addition, while some yards had separate inflow and outflow access for trains, others were terminating yards, so terminating locomotives needed to be taken off the front of trains, and run round to form a new service. Yards had different topography and gradients which necessitated different configurations of handbrakes on sets of wagons. In general, each yard had local idiosyncrasies that shaped (and typically increased) the number and complexity of train moves required. Little or no data was available on the number of moves within a yard.

These moves required not only complex communications between the groundstaff and shunting drivers. Workshop participants noted the reliance on communications and the potential for overfamiliarity leading to deviation from appropriate communication. Furthermore, shunting moves that required trains to leave the confines of the yard and to enter mainline network (albeit sometimes for a matter of a few hundred metres before reversing back into the yard on a different line) required communications with the Network Rail signaller, which was also time constrained by requiring a sufficient gap in other services on the network to accommodate the move.

Staffing varied across yards – some being staffed 24/7 while others were staffed on an occasional basis or operated by staff driving to the yard on an as needed basis. Even when yards were staffed, the supervision (the planning and sequencing of work) may be remote, and this sometimes meant supervisors did not always have a good understanding of local constraints on the ground.

The arrangement of assets and wagons in the yard was also complex. Wagons come in many different types and train sets would often require specific combinations, thus increasing the number of shunt moves as a number of wagons would need to be pulled out from within a larger set of wagons. This complex compilation of wagons also applied to wagons being located and moved to maintenance areas, either for repair or regular inspection.

In terms of the physical environment, walking conditions, lighting, the exposed nature of the yard (with many activities taking place in the open air) all added to the challenges of the work. Several yards are immediately next to live running mainlines. Yards were often broken into two or more separate areas which required walking and sometimes travel by van. This was also additional time to be factored into tasks.

Also, the tasks in the yard were often physically demanding. For example, handbrakes required a high degree of torque, particularly if clogged with debris from travel, or stiff after a period of non-use. Handbrakes could be quite low down on the wagon (e.g. around 0.6) metres and require significant stoop for taller members of staff. Other physically demanding tasks involved removal of stanchions, required to hold loads in place, weighing 7 or 8kg but with several on a wagon (and up to 40 wagons) this could be a significant task. Other tasks included loosening and tightening of couplings between trains, involving both a physical force to perform the work, and a stoop to get under buffers. All of these tasks took time and increased the physical demands of work in the yard.

Figure 1 Key tasks identified in observations and workshop

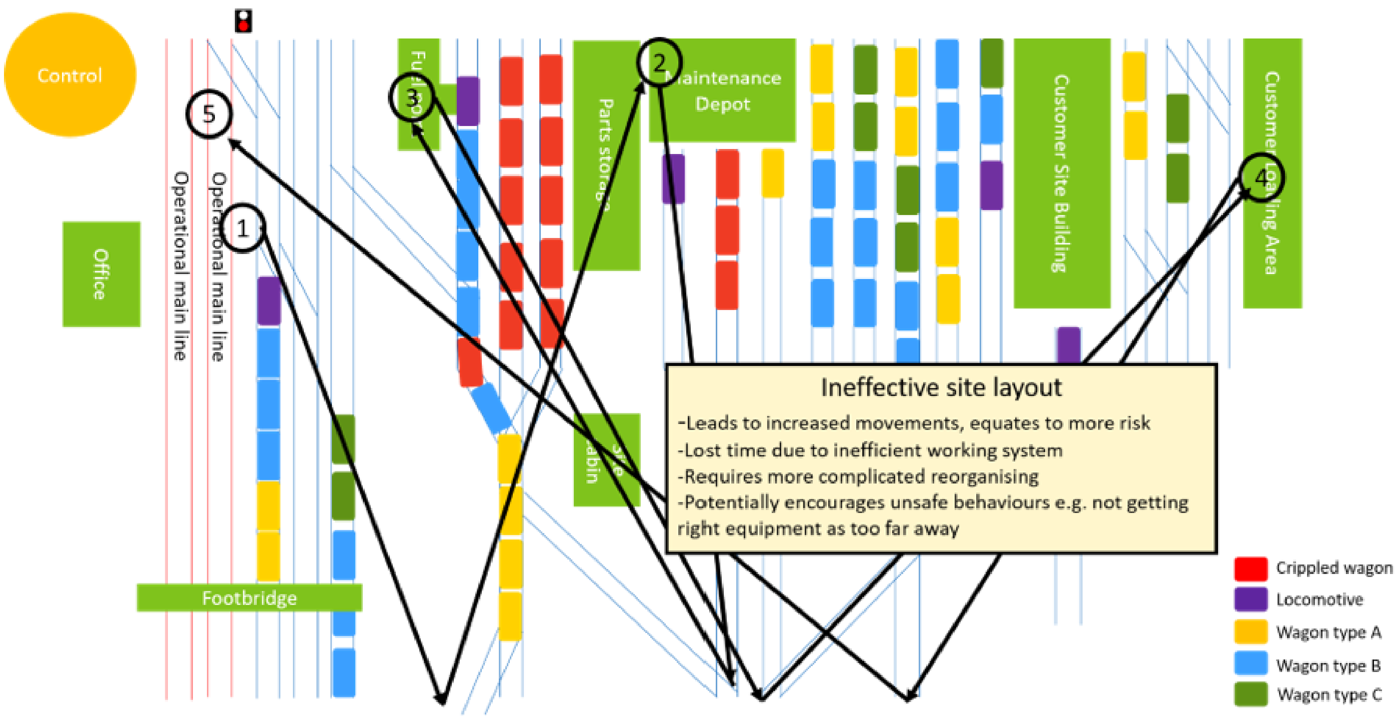


Figure 2 - Schematic of task flows in a generic freight yard

Task Models

Figure 1 shows the high level task model and task flows for work. A key finding from observations and discussion was to consider the relevance of office and planning work feeding into the yard. This involved planning the commercial arrangements, planning of paths and rosters. This proved to be a crucial factor in setting up the task loads in the yard.

These tasks were also subject to their own ergonomics issues, and planning and rostering was still a somewhat manual task that could be prone to errors in data entry, and requiring multiple rounds of checking. Also, there was comment in the workshop that planning and commercial processes often did not take into account the capacity of the yard and practical constraints.

Figure 2 gives a pictorial representation of a generic, large yard. Trains arriving from the mainline (1) may need maintenance (2) and then refuelling (3) before travelling to the loading area to pick up a set of loaded wagons (4) and then onto the network (5). Note how each move requires going out of the yard and then reversing or running the loco around to the front. Even this, however, is an oversimplification, both as observed and verified in the workshop. While the models in Figure 1, and the scenario in Figure 2, are linear, this is not how work is performed. In order to complete tasks as efficiently as possible, multiple tasks would be combined and conducted in parallel. For example, groundstaff might conduct multiple train preparation tasks, or combine together multiple wagons into a single shunt move to cut down on time and best manage the constraints of the freight yard. Also, these timelines belie how much plans are subject to change. Plans would often have to be adapted within short (less than 24 hour) timeframes. A final comment from workshop participants on the task models was that there was insufficient emphasis on the interactions with 3rd parties such as maintainers, 3rd party companies responsible for wagon loading, Network Rail, customers etc. Finally, Table 1 provides a summary of the tactical and strategic solutions to proposed by workshop participants to address human factors issues in the yard.

Discussion

When considered in combination, the freight yard reveals a complex picture of fluid cognitive planning and replanning. The physically demanding nature of the job, plus the need to work around site constraints, further influenced the planning and execution of work. In order to manage the external pressures for delivering freight (subject to short-term replanning), supervisors work with groundstaff to tie multiple tasks together, conducting them in parallel or compiling together for efficiency, to put capacity back into a stretched system. Work in the yard requires a high degree of flexibility, tacit knowledge and cooperation. Space in this paper prevents a full listing of the myriad factors that make work complex in the freight yard.

The work correlates with the limited previous findings so far. First, the understanding of work is difficult through paperwork alone (Bowler and Basicik, 2015) and the difference between work-as-imagined and work-as-done is significant. This is vital as we look towards the introduction of European Train Control System or digital coupling. While such developments may offer key benefits for the freight sector, they cannot be successfully introduced until they fully reflect the complexities of work in the freight yard, and the realities of human factors in the freight environment (Vaghi et al., 2018).

Many of the situational factors such found by Lawton (1998) relevant to the shunting task are found to more widely impact work across the freight yard. What was unexpected was the degree to which back office, planning and commercial processes set the scene and constraints under which freight

Table 1: Workshop participant solutions to human factors

Tactical- Unobtrusive monitoring of comms/actions through CCTV, audio recording, body cams etc
Tactical- Better understanding of freight system in training e.g. planners spending more time with groundstaff to see what role is like.
System- Industry take a lead on tools to make process easier e.g. automatic reading of locos and wagons, recognise where they are in the yard (geo-location)
System- Operationalising a fair culture that applies to staff and senior management. Something similar to Network Rail model and freight life saving rules suggested.
System- Better integration of non-technical skills
System- Develop SSOWs with human factors in mind.
System- Standardised training school for staff- drive a common standard throughout the industry.

yard work operates. Several groundstaff commented that customers drove the demands that need to be followed. In a manner, the freight yard provides the resilience in the wider freight system – this is the point in the network that can handle short-term changes, turn trains around quickly, and adapt to changing customer demands. This is only achievable through the commitment flexibility and adaptability of the workforce. In Woods (2015) definitions of resilience, this is robustness – an ability to soak up, changes and work fluidly, but not necessarily without cost. While the commitment, quality / safety of work and professionalism of on-site staff was evident and paramount, this kind of flexible and adaptive working will inevitably lead to trade-offs and the kind of events identified in Golightly et al (2022).

There are a number of limitations of the work. First, it is primarily wagon-based and, as noted in the workshop, a different process would be observed for the management of dangerous goods. Second, the observations focussed on large yards with on-site or nearby supervision. However, many sites are smaller and operate with remote, mobile working as it is needed. Not only does this work need to be observed, better statistical analysis is required to understand the risk associated with these sites. While they have fewer train movements they may generate disproportionate risk. Finally, the type of analysis we performed was observational and more about the working context. It would be valuable to conduct a more structured cognitive task analysis of the supervisory role in sites. Taking our lead from recent work in areas such as healthcare (e.g. Sanford et al., 2022) this would help us to identify adaptive behaviours that are occurring to balance work within constraints, thus identifying when workarounds are highlighting specific issues, and potential areas for improvement.

Conclusions

Overall, the freight yard has received little research attention, yet is a potential source of risk both to staff working there, and for trains that then head out onto the network. The observational and workshop activities presented in this paper captures the role of people in the freight yard, a unique and challenging environment, highlighting physical risks, but also highly fluid and cognitive planning to achieve success. This work contributes insight to anyone looking at human performance in freight and logistics, and will also be specific relevance to those looking at digital technologies such as ETCS and digital coupling, giving insight into the practicalities of how 'work as done' could impact the acceptability of deployments.

Acknowledgements

First author was funded by Newcastle University QR Supporting Policy Fund.

References

Bowler, N & Basicik, D (2015) A human factors review of the rail operation at the port of Felixstowe. Proceedings for 5th International Rail Human Factors Conference, 2015.

Golightly, D., Lonergan, J., Ethell, D., Gibson, H., & Hill, D. (2022). A structured incident analysis of human performance in freight train preparation. Proceedings of International Rail Safety Council.Seveille, Spain, 2022.

Hricová, R. (2016). RFID as a Tool of Competitiveness Increase of Rail Freight. *International Scientific Journal about Technologies*, *2*(1), 11-14.

Lawton, R. (1998). Not working to rule: understanding procedural violations at work. *Safety science*, *28*(2), 77-95.

Rail Delivery Group (2021) Assessing the Value of Rail Freight. Accessed from https://www.raildeliverygroup.com/about-us/publications/12839-2021-04-assessing-the-value-of-rail-freight/file.html

Rail Accident Investigation Branch (RAIB) (2022) Derailment and fire involving a tanker train at Llangennech, Carmarthenshire 26 August 2020. Available from https://www.gov.uk/government/news/report-012022-derailment-and-fire-involvinga-tankertrain-at-llangennech-carmarthenshire

Ryan, B., Golightly, D., Pickup, L., Reinartz, S., Atkinson, S., & Dadashi, N. (2021). Human functions in safety-developing a framework of goals, human functions and safety relevant activities for railway socio-technical systems. Safety science, 140, 105279.

Sanford, N., Lavelle, M., Markiewicz, O., Reedy, G., Rafferty, A. M., Darzi, A., & Anderson, J. E. (2022). Capturing challenges and trade-offs in healthcare work using the pressures diagram: An ethnographic study. Applied Ergonomics, 101, 103688.

UNESCAP (2021) Enhancing shift towards Sustainable Freight Transport in Asia and the Pacific Opportunities through railway decarbonisation Available from https://www.unescap.org/kp/2021/enhancingshift-towards-sustainable-freighttransport-asia-and-pacific-opportunities

Vaghi, C., Wheat, P., Österle, I., Milottia, A., & Nellthorp, J. (2016). The role of human factors in rail freight innovation. Towards Innovative Freight and Logistics, 2, 245-258.

Woods, D. D. (2015). Four concepts for resilience and the implications for the future of resilience engineering. *Reliability Engineering & System Safety*, *141*, 5-9.

Zhang, Z., Turla, T., & Liu, X. (2021). Analysis of human-factor-caused freight train accidents in the United States. *Journal of Transportation Safety & Security*, *13*(10), 1157-1186.

Responding to rudeness: does instigator status and directness matter?

Amy Irwin[1], Helen Silver-MacMahon[2], Liz Mossop[3], Kendyl Macconnell[1] & Luiz Santos[4]

[1]University of Aberdeen, UK [2]VetLed, UK [3]University of Lincoln, UK [4]University of Glasgow, UK

SUMMARY

The current paper applies a biobehavioural model of workplace incivility to explore responses to rudeness within the veterinary context. Veterinarians and veterinary nurses (*n*=132) were asked to evaluate six fictional scenarios depicting two types of rudeness (direct versus indirect) across three instigators (clients, co-workers and senior colleagues). The findings indicated that direct rudeness (demeaning comments) was appraised more negatively than indirect rudeness (ignoring person). Responses varied across scenarios; direct rudeness was associated with reciprocation, exit, avoidance, discussion with manager and support seeking, whereas indirect rudeness was associated with affiliative and ignoring responses. There was a significant positive association between appraisal and confrontation, exit, avoidance, support seeking and reporting responses. The findings confirm the utility of the biobehavioural model of incivility response and build on this model in terms of variation in response selection according to directness and status effects. From a practical perspective the findings suggest that interventions to manage rudeness in veterinary practice should accommodate variation in rude behaviours and include tailored responses based on instigator.

KEYWORDS

Incivility, Status, Directness, Veterinary

Introduction

'It totally affects your day because you start to question was it me? Was it something I did? Is it my professionalism?' (Vet describing the impact of client rudeness; Irwin, Hall & Ellis, 2022a).

Workplace mistreatment is a broad concept, encompassing aggression, harassment, ostracism and incivility, all of which can have adverse consequences (including reduced wellbeing, job satisfaction and work performance, Yao, Lim, Guo et al., 2022) for the worker experiencing these behaviours. Within that broad umbrella, incivility can be distinguished from other forms of mistreatment via three key mechanisms; uncivil behaviours are perceived as minor, or low-level, non-physical acts (an example of an uncivil behaviour could be scowling at somebody, in contrast to an aggressive act such as physical intimidation); incivility can be ambiguous in terms of intent to harm (e.g. the uncivil act of not responding to someone during a meeting may be due to the instigator being distracted, in contrast mistreatment via undermining an employee to reduce success is clearly intentional) (Yao et al., 2022). Thirdly, incivility goes against workplace norms for appropriate or polite behaviour, meaning perception of incivility can vary across contexts and individuals (Andersson & Pearson, 1999). Incivility can be either active / direct, described as a commission of disrespect (e.g. unpleasant comments, sarcasm), or passive / indirect, involving omission of respect (e.g. ignoring a request via email) (Yuan, Park & Sliter, 2020).

A recent meta-analysis reported incivility as a reliable, valid construct with impacts independent of other types of mistreatment, highlighting the need for a tailored approach to addressing uncivil behaviours in the workplace (Yao et al., 2022). This is particularly important given the relatively subtle nature of incivility, increasing the risk of such behaviours being ignored at an organisational level, despite the associated negative ramifications (Cortina et al., 2017). Incivility appears inherent in the workplace, with approximately 98% of employees estimated to experience some form of incivility at work and 50% experiencing incivility at least once a week (Porath & Pearson, 2012). Despite this, there is relatively little research about the coping strategies used in response to incivility, or the extent to which these strategies are considered appropriate and / or effective across different job roles. The aim of the current study was to explore the likelihood of different responses to direct and indirect rudeness across three instigator types; client, co-worker and senior colleague, within the specific context of veterinary practice.

Incivility in veterinary practice

Veterinary staff are acknowledged as having a stressful occupation, one with high levels of suicidal ideation and burnout (Andela, 2020). Research indicates a range of psychosocial stressors within this environment, including long work hours, financial insecurity and, most relevant to the current paper, negative interactions with clients and co-workers, including management of unrealistic client expectations, and conflict with colleagues (Bartrum et al., 2009). More specifically, experiencing incivility from clients and co-workers can have an adverse impact on job satisfaction, and mental health, as well as increasing quitting intention and the risk of burnout of veterinary staff (Irwin, Silver-MacMahon & Wilcke, 2022b). The range of interactions necessary for veterinary work, with both clients and colleagues, combined with the potentially harmful impact of incivility, emphasise the need for further work examining incivility coping responses within this context.

Status

Cognitive appraisal theory describes incivility via a three-step process whereby the target assesses the situation to determine threat level and select the most appropriate response (Cortina & Magley, 2012). This encompasses appraisal of the potential for negative impact, consideration of potential responses and coping strategies, and evaluation of the potential interaction outcome (Cortina & Mageley, 2012). Social power, or where the instigator and target sit within the organisational hierarchy, has been suggested as a factor within this assessment. For example, Porath and Pearson (2012) report that targets of incivility who evaluate their status as higher than the instigator tend to react aggressively, whereas lower status victims may be more likely to withdraw. Demographic characteristics can influence both vulnerability to incivility, and the selection of a response or coping strategy. Specifically, women appear more likely to exit the interaction, and men more likely to respond aggressively (Cortina & Mageley, 2009). Within the veterinary context research indicates that veterinary nurses experience higher levels of incivility than veterinarians, with an associated risk of burnout as a result (Irwin et al., 2022b). Appraisal theory highlights the importance of understanding how individuals appraise and understand uncivil behaviours according to status – of both the instigator and the victim, since this will also impact their responses and the potential consequences of experiencing the behaviour.

Coping

Coping encompasses any actions or thought processes used to manage stressful situations. Targets of workplace incivility can utilise a range of coping responses, which differ across and within individuals (Cortina & Mageley, 2003). Passive strategies (such as conflict avoidance) tend to be used more frequently than active coping strategies (such as confrontation), with researchers suggesting this may be linked to the difficulties inherent in reporting a low-level and potentially ambiguous behaviour (Cortina & Mageley, 2009). Research suggests that both passive and active strategies may be ineffective in preventing future incivility, but that active strategies may support psychological forgiveness, helping the target move on from the incident (Hershcovis et al., 2018).

Previous research with veterinary staff highlights a preference for utilising social support, and attempting to ignore uncivil behaviours, along with emphasising the importance of a supportive practice culture (Irwin et al., 2022a).

Cortina and colleagues (2021) recently produced a biobehavioural model of workplace incivility, which suggests that uncivil acts prompt appraisal, and a biological response from the victim. This biological reaction leads to a behaviour response, with potential options categorised across four quadrants: reciprocation, retreat, relationship repair and recruitment of support. Reciprocation and relationship repair both involve direct efforts to change the behaviour of the instigator, whereas retreat and recruitment of support encompass leaving the immediate situation. Moreover, reciprocation and retreat both reduce social connections and are likely to prolong the initial biological response (e.g. heightened adrenalin), whereas relationship repair and recruitment of support both involve increasing social connections (with the instigator or with others) and reducing the original biological response. The model seeks to advance study of incivility responses and provide avenues of exploration in terms of the advantages and disadvantages of each response type. The current paper applies this model to responses to incivility within the veterinary context, and across different instigators and forms of rudeness, in order to further our understanding of response options and coping mechanisms.

Study aims

The aim of the current vignette study was to investigate perception, appraisal and response to incivility within veterinary practice across multiple sources, rudeness type and job role.

Method

Participants

A total of 132 participants (111 female, mean age: 38.1yrs, mean years job experience: 7.4yrs) were recruited via social media and direct email invitation. The sample comprised of veterinary surgeons (n = 76), veterinary nurses (n = 53), and not stated (n = 3).

Questionnaire

The online questionnaire consisted of two main sections and was created using SNAP. The first section comprised questions relevant to demographic information, including job role, years of job experience, nationality, gender, age, practice status and practice focus.

The second section encompassed six vignettes, followed by a series of scales and items relevant to rudeness, appraisal and response. The scenarios were designed according to the experimental vignette method (Aguinis & Bradley, 2014), whereby two main variables were manipulated: directness (indirect versus direct rudeness) and instigator status (client, co-worker, supervisor). The vignettes were drawn from previous research exploring rudeness experiences in veterinary practice (Irwin et al., 2022a; 2022b) and were checked by veterinary experts to ensure relevance and realism.

The vignettes manipulated the directness of the rudeness shown, with indirect incivility depicted as the instigator ignoring the victim, and direct incivility shown as demeaning comments related to work performance. The scenarios were created to show incivility originating with three different instigators: clients, co-workers or senior colleagues. The vignettes were presented in a randomised order using the randomise function within SNAP software. Following each vignette there were a series of quantitative items designed to investigate perceived rudeness level, appraisal and coping response. First, participants were presented with a single item: *If this behaviour happened in real-life would you consider the behaviour shown here to be rude?* with five response options from 1 –

not at all rude, to 5 – extremely rude. This was followed by a short scale designed to assess participant's appraisal of the depicted behaviour. The scale asks respondents to characterise their perception of behaviour: *If this situation occurred in real-life to what degree would you find it:* followed by six descriptors (e.g. offensive, annoying, frustrating) with five response options (from 1 – not at all, to 5 – extremely) (Cortina & Mageley, 2009). Next, participants were asked: *If this situation occurred in real-life how likely would you be to respond in the following ways?* This was followed by eight items, five of which were drawn from the Coping with Harassment questionnaire (CHQ, Cortina & Mageley, 2009) (e.g. 'ignore it', 'let the person know you didn't like their behaviour') together with three original items designed to reflect response options from the biobehavioural theory of response to workplace incivility (Cortina, Hershcovis & Clancy, 2021) which were not reflected within the CHQ (e.g. 'make a friendly overture').

Results

The mean score for each dependent variable was calculated (Table 1) across the six vignettes in order to gain an overview of the general pattern of results. These preliminary findings indicate that responses varied across the vignettes, with a general overview suggesting that direct rudeness tended to be appraised more negatively than indirect rudeness, though it was not always perceived as more rude.

Table 1: Mean scores (standard deviation) for perceived rudeness, appraisal and initial reported response to rudeness vignettes.

Variable	**Client**		**Co-worker**		**Senior**	
	Indirect	Direct	Indirect	Direct	Indirect	Direct
Rudeness level	4.08 (.86)	4.08 (.87)	3.26 (1.02)	4.27 (.76)	3.62 (1.04)	3.13 (1.05)
Appraisal	16.51 (4.45)	20.36 (5.20)	14.35 (4.38)	20.71 (5.23)	16.12 (5.20)	19.14 (5.25)
Friendly overture *(affiliative)*	2.78 (1.41)	2.67 (1.38)	2.38 (1.32)	2.00 (1.29)	2.84 (1.37)	2.15 (1.36)
Ignore it *(ignore)*	2.31 (1.28)	1.73 (1.10)	2.61 (1.32)	1.91 (1.06)	2.77 (1.22)	1.53 (1.07)
Let person know you didn't like their behaviour *(confront)*	2.82 (1.43)	2.70 (1.33)	2.40 (1.37)	3.45 (1.33)	2.19 (1.24)	2.13 (1.29)
Reciprocate *(reciprocate)*	1.11 (.34)	1.19 (.50)	1.17 (.56)	1.69 (1.17)	1.15 (.46)	1.19 (.55)
Leave situation *(exit)*	1.59 (1.11)	2.27 (1.35)	2.66 (1.39)	2.93 (1.41)	2.43 (1.25)	2.06 (1.30)
Try to avoid that person *(avoid)*	1.56 (.93)	2.91 (1.38)	1.65 (.95)	2.68 (1.44)	2.17 (1.28)	2.35 (1.33)
Talk to friend / family member *(support)*	2.93 (1.50)	3.65 (1.36)	2.49 (1.47)	3.58 (1.37)	3.02 (1.50)	3.88 (1.24)
Talk to senior colleague / manager *(discuss)*	2.64 (1.43)	3.94 (1.17)	1.99 (1.26)	3.38 (1.42)	2.33 (1.32)	2.97 (1.43)

Impact of directness and instigator status on reported rudeness level, appraisal and response.

A series of within subjects ANOVA (2 x directness, 3 x status) analyses were used to explore whether responses varied across vignettes ($p<.005$ was set as the significance level to ensure rigour when conducting multiple analyses). Results for each dependent variable can be viewed within Table 2.

Table 2: Within subjects ANOVA (2 x directness, 3 x status) for level of rudeness, appraisal and response.

Variable	Mean (SE)	Directness	Mean (SE)	Status of instigator	Interaction
Rudeness level	Indirect: 3.65 (.066) Direct: 3.84 (.061)	$F(1, 125)$: 7.853, p: .006	Client: 4.08 (.066) Co-worker: 3.76 (.060) Senior: 3.38 (.073)	$F(2, 250)$: 52.338, $p<.001$	$F(2, 250)$: 72.754, $p<.001$
Appraisal	Indirect: 15.79 (.36) Direct: 20.14 (.39)	$F(1, 119)$: 205.258, $p<.001$	Client: 18.58 (.41) Co-worker: 17.56 (.37) Senior: 17.75 (.39)	$F(2, 238)$: 5.732, p: .004	$F(2, 238)$: 18.730, $p<.001$
Affiliative	Indirect: 2.64 (.10) Direct: 2.26 (.09)	$F(1, 123)$: 18.518, $p<.001$	Client: 2.71 (.11) Co-worker: 2.15 (.09) Senior: 2.28 (.10)	$F(2, 246)$ 18.007, $p<.001$	$F(2, 246)$: 6.201, p: .002
Ignore	Indirect: 2.57 (.09) Direct: 1.72 (.07)	$F(1, 120)$: 85.497, $p<.001$	Client: 2.02 (.09) Co-worker: 2.26 (.09) Senior: 2.15 (.08)	$F(2, 240)$: 2.945, p:.055	$F(2, 240)$: 8.110, $p<.001$
Confront	Indirect: 2.47 (.10) Direct: 2.76 (.09)	$F(1, 120)$: 12.910, $p<.001$	Client: 2.76 (.11) Co-worker: 2.92 (.10) Senior: 2.16 (.10)	$F(2, 240)$: 32.984, $p<.001$	$F(2, 240)$: 23.949, $p<.001$
Reciprocate	Indirect: 1.15 (.03) Direct: 1.36 (.05)	$F(1, 122)$: 19.044, $p<.001$	Client: 1.15 (.03) Co-worker: 1.43 (.06) Senior: 1.17 (.04)	$F(2, 244)$: 17.698, $p<.001$	$F(2, 244)$: 14.858, $p<.001$
Exit	Indirect: 2.24 (.08) Direct: 2.42 (.09)	$F(1, 121)$: 3.728, p:.056	Client: 1.93 (.09) Co-worker: 2.79 (.10) Senior: 2.25 (.09)	$F(2, 242)$: 36.323, $p<.001$	$F(2, 242)$: 15.638, $p<.001$
Avoid	Indirect: 1.78 (.08) Direct: 2.64 (.11)	$F(1, 119)$: 108.151, $p<.001$	Client: 2.25 (.09) Co-worker: 2.11 (.09) Senior: 2.26 (.10)	$F(2, 238)$: 2.037, p:.133	$F(2, 238)$: 31.672, $p<.001$
Support	Indirect: 2.81 (.12) Direct: 3.69 (.11)	$F(1, 122)$: 126.796, $p<.001$	Client: 3.28 (.12) Co-worker: 3.02 (.12) Senior: 3.46 (.11)	$F(2, 244)$: 14.343, $p<.001$	$F(2, 244)$: 2.532, p:.082
Discuss	Indirect: 2.33 (.09) Direct: 3.43 (.09)	$F(1, 122)$: 199.818, $p<.001$	Client: 3.31 (.10) Co-worker: 2.69 (.10) Senior: 2.64 (.11)	$F(2, 244)$: 28.137, $p<.001$	$F(2, 244)$: 8.111, $p<.001$

* Grey squares denote non-significant results.

The findings indicate a consistent impact of directness on reported responses (excluding rudeness level and the exit responses), whereby direct rudeness was appraised more negatively than indirect rudeness and more likely to lead to a confront, reciprocation or avoidance response. This is with the caveat that the indicated likelihood for reciprocation was low (<2) across all vignettes, suggesting that reciprocation as a response is not very likely in general within veterinary practice regardless of rudeness type. Direct rudeness was also more likely to prompt support seeking and discussion than indirect rudeness (Table 2). In contrast, indirect rudeness was linked to a greater likelihood of ignoring the behaviour, The impact of status was more variable and often qualified by a significant interaction. For example, there was a significant main effect status, with a significant interaction, for making an affiliative (friendly) response (Table 2). The interaction (Table 1) indicates that although the likelihood of a friendly gesture was less likely in response to direct, versus indirect, rudeness across all three status types, this difference was least pronounced across the client vignettes; suggesting a friendly gesture is a potential response to client rudeness regardless of directness. The support response was the one exception to this, with a significant main effect of directness and status, with no significant interaction (Table 2). The pattern of results (Table 1) indicates that talking to a friend was more likely in response to direct versus indirect rudeness across all three status types. In addition, talking to a friend was more likely in response to senior colleague, than client rudeness, and more likely in response to both of those status types than co-worker rudeness.

Impact of job role on reported rudeness level, appraisal and response to incivility vignettes.

A series of one-way ANOVA's (used rather than t-tests to reduce the likelihood of type 1 error) were conducted to compare responses across job roles (veterinarian versus veterinary nurse) for all six vignettes.

There were no significant differences across job role for perception of rudeness, appraisal or any of the response types excluding one significant difference for the vignette depicting indirect rudeness from a senior staff member. For that vignette there was a significant difference in the reported likelihood of letting the person know you didn't like their behaviour, with veterinarians less likely (M: 1.92, sd: 1.1), than veterinary nurses (M: 2.65, sd: 1.35) to report this response ($F(1, 122)$: 11.010, p:.001).

Associations between variables

A correlation matrix was developed to examine the associations between the key dependent variables across the six vignette conditions (Table 3). The results indicate a consistent positive association between rudeness level and negative emotional reaction across all six scenario. There are also fairly consistent positive relations between rudeness level, appraisal and likelihood of confrontation, exit, avoidance, support seeking and reporting (with some variability across scenario types). There was a lack of a consistent relationship between rudeness, appraisal and affiliative, ignore or reciprocate responses.

Table 3: Pearson correlations illustrating associations between rudeness level, appraisal and responses across vignette conditions.

Vignette	**Rude/ appraise**	**App-raise**	**affiliative**	**ignore**	**confront**	**Recipr-ocate**	**exit**	**avoid**	**support**	**report**
Client indir	Rudeness	.620**	.010	.128	.236*	.150	.251**	.203*	.312**	.263**
	Appraisal		-.021	.097	.163	.185*	.267**	.300**	.370**	.395**
Co-worker indir	Rudeness	.670**	-.011	-.302**	.367**	-.110	.214*	.241**	.222*	.463**
	Appraisal		-.004	-.132	.260**	-.178	.281**	.482**	.282**	.281**
Senior indir	Rudeness	.791**	-.011	-.178	.311**	.044	.399**	.330**	.495**	.359**
	Appraisal		.022	-.048	.220*	.092	.494**	.449**	.456**	.291**
Client direct	Rudeness	.644**	-.182*	-.052	.457**	.202*	.272**	.216*	.336**	.183*
	Appraisal		-.186*	-.013	.308**	.235*	.254**	.354**	.402**	.288**
Co-worker direct	Rudeness	.693**	.053	.131	.235*	.039	.111	.133	.326**	.292**
	Appraisal		.016	.101	.062	.126	.219*	.317**	.382**	.208*
	Rudeness	.664**	.046	-.023	.369**	.167	.407**	.433**	.244**	.167

Senior direct	Appraisal		.009	.017	.250**	.248**	.479**	.572**	.366**	.197*

**significant at <.005, * significant at <.05, grey squares denote non-significant correlations

Discussion

The analysis reported above indicates that the selection of response will vary according to the status of the instigator and the nature of the rude behaviour. Direct rudeness, in this case demeaning comments, was appraised more negatively (e.g. considered more offensive) than indirect rudeness, and prompted an increased reported likelihood of the victim responding by being rude themselves, confronting the behaviour, exiting the situation, avoiding the instigator in the future, seeking support from friends and discussing the behaviour with a senior colleague. Indirect rudeness, in this case ignoring someone, was associated with an increased likelihood of a friendly response (e.g. a smile) or ignoring the behaviour. Status effects were quite variable, and qualified by interactions, but general trends suggest that a friendly gesture was more likely in response to client, as opposed to co-worker or senior colleague rudeness. Seeking support was most likely in response to direct rudeness from a senior colleague, and discussing the incident with a senior colleague was most likely in response to direct client rudeness. The reported responses to the vignettes did not vary significantly across job roles. Finally, the correlation matrix indicates that negative appraisals of rudeness are associated with confrontation, exit, avoidance, support seeking and reporting responses.

The results build on the biobehavioural theory put forward by Cortina and colleagues (2021) by suggesting that selection of an appropriate response to rude behaviour at work can be influenced by the status of the person engaging in the behaviour, as well as the nature of the behaviour itself. Based on their theory of biological response, it is possible that direct rudeness prompts a heightened biological reaction in comparison to indirect rudeness (a suggestion supported by the significantly more negative appraisals of direct rudeness, and the association between rudeness level and appraisal), increasing the potential likelihood of the associated 'fight or flight' response, and producing behaviours associated with reciprocation (fight) and retreat (flight). This heightened response may also explain the increased likelihood of seeking support, which can be interpreted as seeking out comfort and support, which may soothe and reduce the biological reaction. Indirect rudeness, linked to a reduced biological reaction, enables relationship repair to be attempted via a friendly overture. A lower biological reaction may also make ignoring the behaviour easier.

Within the incivility literature, appraisal of uncivil behaviours has consistently been highlighted as an important aspect of understanding incivility within the workplace. Researchers emphasise that rudeness can produce an emotional reaction within the target, but that this emotional appraisal should be relatively mild to match the low intensity of rudeness as a mistreatment construct (Cortine & Mageley, 2009). The current results suggest that not all rude behaviours should be considered equal, with direct rudeness prompting more negative appraisals (indicating a higher level of frustration, annoyance, upset etc.) than indirect rudeness. As such organisational interventions may need to encompass guidance about the variety of rude behaviours that might be experienced, along with tailored support mechanisms according to directness.

Veterinary practice is based on successful interactions with clients, as such it is perhaps not surprising that participants within the current study showed a preference for making a friendly overture when the instigator was a client. This may be partially explained through the mechanism of emotional labour, where workers within service industries – reliant on client business – are expected to present positive emotions (service with a smile) to clients (Yagil, 2021). Such overtures could be potentially harmful to mental health, with research indicating that suppression of negative emotions, particularly when combined with portraying false positive emotions, can reduce employee wellbeing (Goldberg et al., 2007; Irwin et al., 2022b). However, this type of affiliative response might also be aimed at reducing the cause of the instigator's rudeness, and as such may comprise a practical approach to managing client emotions (Irwin et al., 2022a). Similarly, ignoring rude behaviour from clients has been discussed within previous research as a mechanism for maintaining professionalism and maintaining a calm façade (Irwin et al., 2022a). It is important to raise awareness of the effort involved in such mechanisms, with support provided for veterinary staff following such interactions.

The findings confirm support seeking as a popular response to the experience of incivility, regardless of the status of the instigator. Support seeking has been previously highlighted as a frequently used response by veterinary staff (Irwin et al., 2022a; 2022b), enabling the victim of the behaviour to 'let off steam' by discussing the incident with colleagues, as well as gaining empathy and the benefit of insight into shared experiences. Although discussion with a senior colleague might produce many of the same benefits, this response appeared more frequent when dealing with client as opposed to colleague rudeness. It is unclear from the current results why this might be the case but may relate to difficulties in reporting a relatively low-level behaviour using official channels highlighted by Cortina and Mageley (2009). However, given the emphasis on leader commitment to professional behaviour as a mechanism for managing unprofessional behaviours (including incivility) within healthcare research (Hickson et al., 2007), this aspect of responding to incivility may benefit from further research.

Limitations

The current study is based on self-reported responses to fictional scenarios, and as such may not represent the full range of responses seen in real-world practice. In addition, this data is primarily subjective, and as such future research should look at gathering objective data (such as physiological reactions to rudeness and coping mechanisms) to validate this pattern of results. Finally, the sample is representative of veterinary staff from the UK and Ireland and as such the results may not generalise beyond that population and geographical location.

Conclusion

The findings confirm the utility of the biobehavioural model of incivility response and builds on this model to point to variation in response selection according to the directness of rudeness shown, and the status of the instigator. From a practical perspective the findings suggest that interventions to manage rudeness in veterinary practice should accommodate variation in rude behaviours and include tailored responses based on instigator. In addition, it is important to recognise the value of support for victims of rude behaviour, particularly where the victim may have had to manage their emotional response.

References

Aguinis, H. & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods, 17*, 351-371.

Andela M. (2020). Burnout, somatic complaints, and suicidal ideations among veterinarians: Development and validation of the Veterinarians Stressors Inventory. *Journal of Veterinary Behavior*, *37,* 48-55.

Andersson L.M., Pearson C.M. (1999). Tit for tat? The spiralling effect of incivility in the workplace. *Academy of Management Review*, *24,* 452-71

Bartram D.J., Yadegarfar G., Baldwin D.S. (2009). Psychosocial working conditions and work-related stressors among UK veterinary surgeons. *Occupational Medicine*, *59,* 334-341.

Cortina, L. M., Hershcovis, M., & Clancy, K. B. (2022). The embodiment of insult: A theory of biobehavioral response to workplace incivility. *Journal of Management*, *48*, 738-763.

Cortina L.M., Kabat-Farr D., Magley V.J. & Nelson, K. (2017). Researching rudeness: The past, present, and future of the science of incivility. *Journal of Occupational Health Psychology*, *22,* 299-321.

Cortina L.M. & Magley, V.J. (2003). Raising voice, risking retaliation: Events following interpersonal mistreatment in the workplace. *Journal of Occupational Health Psychology, 8*, 247-252.

Cortina, L.M. & Magley, V.J. (2009). Patterns and profiles of response to incivility in the workplace. *Journal of Occupational Health Psychology, 14*, 272-283

Goldberg L.S., Grandey A.A. (2007). Display rules versus display autonomy: emotion regulation, emotional exhaustion, and task performance in a call center simulation. *Journal of Occupational Health Psychology*, *12,* 301-310.

Hershcovis M.S., Cameron A.F., Gervais L., Bozeman J. (2018). The effects of confrontation and avoidance coping in response to workplace incivility. *Journal of Occupational Health Psychology*, *23,* 163-174.

Hickson, G.B., Pichert, J.W., Webb, L.E. & Gabbe, S.G. (2007). A complementary approach to promoting professionalism: identifying, measuring and addressing unprofessional behaviours. *Academic Medicine, 82,* 1040-1048.

Irwin, A., Hall, D. & Ellis, H. (2022a). Ruminating on rudeness: Exploring veterinarian's experiences of client incivility. *Veterinary Record, 190,* e1078.

Irwin, A., Silver-MacMahon, H. & Wilcke, S. (2022b). Coping and consequences: Investigating client, co-worker and senior colleague incivility within veterinary practice. *Veterinary Record, 191,* e2030.

Porath, C. L., & Pearson, C. M. (2012). Emotional and behavioral responses to workplace incivility and the impact of hierarchical status. *Journal of Applied Social Psychology, 42,* 326–357.

Yagil D. (2021). Abuse from organizational outsiders: Customer aggression and incivility. *Special topics and particular occupations, professions and sectors*, *1,* 109-34.

Yao J., Lim S., Guo C.Y., Ou. A.Y. & Ng, J.W.X. (2022). Experienced incivility in the workplace: A meta-analytical review of its construct validity and nomological network. *Journal of Applied Psychology*, *107,* 193-205.

Yuan Z., Park Y., Sliter M.T. (2020). Put you down versus tune you out: Further understanding active and passive e-mail incivility. *Journal of Occupational Health Psychology, 25,* 330-342.

Behavioural intention of e-scooter use: A comparison of users and non-users

İbrahim Öztürk[1] & Nazlı Akay[2]

[1]Institute for Transport Studies, University of Leeds, UK, [2]Department of Psychology, Middle East Technical University, Turkey

SUMMARY

In the present study, the behavioural intention and factors affecting the use of e-scooters were examined among young people in Turkey. Previous users and non-users of e-scooters were compared in relation to the various variables. The relationships of attitudes, subjective norms, perceived behavioural control, and perceived usefulness were investigated regarding the intention to use e-scooters. While users reported positive socio-psychological factors related to e-scooter use, all different factors were positively related to behavioural intention. The results highlighted the importance of socio-psychological factors in predicting behavioural intention, as well as differences between e-scooter users and non-users in these factors.

KEYWORDS

Behavioural intention, e-scooter, micro-mobility

Introduction

Technological developments have enabled road users to own and share new micro-mobility devices, such as electric scooters (e-scooters). The use of e-scooters in traffic has become increasingly common in recent years, having a significant impact on the environment and on how people travel. The introduction of e-scooters, as an alternative mode of transport for short trips, is seen as a mode of transport that is fun, good for the environment, convenient and faster than walking (Sanders et al., 2020).

At these earlier stages, it is crucial to understand road users' intentions towards novel technologies, such as e-scooters, along with the antecedents. Previous studies have shown the importance of socio-psychological factors in influencing road users' acceptance of new technologies and practices, such as automated vehicles (e.g., Buckley et al., 2018; Madigan et al., 2017) and public transport systems (e.g., Chen & Chao, 2011).

To this end, a number of studies have been conducted in recent years to better understand the perception of e-scooter use (e.g., Almannaa et al., 2021; Öztaş Karlı et al., 2022; Ratan et al., 2021; Rejali et al., 2021) through the use of various theories and models such as the theory of planned behaviour (Ajzen, 1991) and the technology acceptance model (Davis, 1989). For example, studies found positive effects of perceived usefulness (e.g., Javadinasr et al., 2022; Ratan et al., 2021; Rejali et al., 2021), perceived ease of use (e.g., Javadinasr et al., 2022; Rejali et al., 2021), and subjective norms/social influence (e.g., Javadinasr et al., 2022; Öztaş Karlı et al., 2022; Rejali et al., 2021) on behavioural intention.

Prior literature has also shown that user/non-user differences are important on several dimensions (e.g., Almannaa et al., 2021; Buehler et al., 2021; Petzoldt et al., 2021; Sanders et al., 2020). Petzoldt et al. (2021), for example, investigated the knowledge of and compliance with rules among users and non-users of electric scooters. Differences in terms of rule knowledge and agreement were observed between e-scooter users and non-users, and also among users. These results indicated the importance of information coming from these groups. Against this background, the present study aimed to understand the factors that influence the intention to use e-scooters and to compare previous users and non-users of e-scooters with respect to socio-psychological factors.

Method

Participants

A total of 443 young people aged between 18 and 25 years (M = 21.25, SD = 1.48, sex: 302 females, 137 males, 4 other) participated in the study. Of the participants, 209 (47.2%) reported having used an e-scooter at least once (age: M = 21.45, SD = 1.42, sex: 127 females, 81 males, 1 other). The remaining 234 participants (52.8%) had never used an e-scooter (age: M = 21.07, SD = 1.50, sex: 175 females, 56 males, 3 other).

Measurements

The survey consisted of several sections. In the first part, questions on demographics, technology acceptance and previous use of e-scooters were included. The second part focused on different aspects of the theory of planned behaviour (Ajzen, 1991) and the technology acceptance model (Davis, 1989) as well as the facilitating conditions. The final item pool was developed based on a review of the literature (e.g., Buckley et al., 2018; Chen & Chao, 2011; Madigan et al., 2017). As a result, attitudes towards e-scooter use were measured using four items (bad-good, stupid-smart, harmful-beneficial, negative-positive) on a 7-point Likert scale with a Cronbach's alpha reliability of .89. A 25-item scale with 5-point Likert scale (from 1: strongly disagree to 5: strongly agree) was developed to measure the remaining constructs.

Procedure

Following the development of the measures for this study, ethical approval was obtained from Middle East Technical University (170 ODTU 2020). In addition to the measurements reported in this paper, the survey also included two questionnaires on personality and values (results not reported here). This study was disseminated using social media and the Sona Systems account of the Department of Psychology, Middle East Technical University. Data were collected by using an online survey platform between October 2020 and June 2021. Participants were provided with an informed consent form prior to entering the survey, and their anonymity and confidentiality were assured. Participants who participated through Sona received bonus points in their courses.

Analyses

The analyses were conducted using SPSS (version 26). First, a principal component analysis with direct Oblimin rotation was performed to examine distinct factorial structures, as the items were obtained from different sources for the study. Following that, a one-way ANOVA was conducted to examine the differences between e-scooter users and non-users in terms of attitudes, perceived behavioural control, perceived usefulness, subjective norms, facilitating conditions, and behavioural intention factors separately. In the final step, a hierarchical regression analysis was conducted to examine the roles of the attitudes, perceived behavioural control, perceived usefulness, subjective norms, and facilitating conditions over behavioural intention after controlling for the effects of sex, age, technology adoption, and previous use of e-scooters.

Results

Behavioural intention towards e-scooter use

The Bartlett's test of sphericity yielded a significant result ($\chi2(300) = 5410.71$, $p < .001$) and the Kaiser-Meyer-Olkin measure of sampling adequacy was .88, indicating that the correlation matrix generated by the items is factorable. Based on the eigenvalues greater than one criterion (Reise et al., 2000), scree plot (Stevens, 2009) and the parallel analysis (O'Connor, 2000), the results supported the five-factor solution. In the final solution, two items were excluded due to factor loadings below the .40 cut-off, and one item was excluded due to not loading on the relevant factor and a decrease in Cronbach's alpha reliability. The total of 22 items explained 64.46% of the variance of the scale. The factor loadings are shown in Table 1.

Table 1. Constructs and factor loadings

Construct	Adapted item	Factor Loading
Perceived behavioural control ($\alpha = .92$)	I can handle an e-scooter with ease.	.905
	I can ride an e-scooter.	.842
	I cannot use an e-scooter.	-.839
	It is easy for me to use an e-scooter.	.831
	I can drive an e-scooter without much mental effort.	.795
	I have the necessary knowledge to use an e-scooter.	.788
	I can use e-scooter applications easily.	.663
Perceived usefulness ($\alpha = .79$)	Using an e-scooter helps me with my transport activities	.768
	Using an e-scooter saves me time.	.695
	There are advantages to using an e-scooter for everyday transport.	.690
	Using an e-scooter is useful in traffic.	.683
	E-scooters are budget-friendly.	.581
	Using e-scooters is good for the environment.	.536
Subjective norms ($\alpha = .71$)	My close circle (e.g., family and friends) is positive about using e-scooters.	.834
	The people around me (e.g., family and friends) is generally supportive of e-scooter use.	.805
	The general public is generally positive about using e-scooters.	.646
Facilitating conditions ($\alpha = .46$)	The transport infrastructure (roads, traffic signals, etc.) is suitable for the use of e-scooters.	.699
	E-scooters are compatible/integrated with other modes of transport I use.	.569
	The media and/or policy makers support the use of e-scooters.	.568
Behavioural intention ($\alpha = .91$)	I would use an e-scooter in the future	-.888
	I plan to use an e-scooter in the near future.	-.885
	I intend to use an e-scooter.	-.774

User and non-user differences

In comparison to e-scooter non-users (Table 2), previous users scored higher on technology adoption, perceived behavioural control, perceived usefulness, and behavioural intention, and had more positive attitudes towards e-scooters. Subjective norms and facilitating conditions did not differ significantly between previous users and non-users.

Table 2. Comparison of users and non-users of e-scooters

Construct	Mean (SD) of users	Mean (SD) of non-users	df	F	p	η_p^2
Technology adoption	3.36 (1.15)	2.47 (1.09)	1,441	69.66	<.001	.14
Attitudes	5.76 (1.21)	5.40 (1.22)	1,441	9.84	.002	.02
Perceived behavioural control	4.25 (.72)	3.43 (.75)	1,441	138.19	<.001	.24
Perceived usefulness*	3.93 (.63)	3.69 (.53)	1,409.38	18.36	<.001	.04
Subjective norms	3.53 (.69)	3.42 (.67)	1,441	2.62	.106	.01
Facilitating conditions	2.62 (.68)	2.57 (.63)	1,441	.62	.431	.01
Behavioural intention	3.72 (.95)	3.18 (1.04)	1,441	32.78	<.001	.07

* Welch statistic was reported.

Determining behavioural intention

The regression analysis (Table 3) focusing on behavioural intention was significant ($F(9, 433) = 40.90$, $p < .001$). Analyses of the first step variables revealed that, females, people with high technology adoption and those who had previously used e-scooters had a higher behavioural intention to use e-scooters in the future. All variables at the second step (i.e., attitudes, subjective norms, perceived behavioural control, perceived usefulness and facilitating conditions) were positively related to behavioural intention, meaning that these aspects, as hypothesised, were associated with the participants' increased behavioural intentions to use e-scooters.

Table 3. Factors associated with behavioural intention

	Behavioural intention					
	R^2	$\triangle R^2$	*df*	*FΔ*	*β*	*p*
1. Step	.18	.18	4,438	23.36		<.001
Sex (female, male)					-.112	.013
Age					.060	.171
Technology adoption					.352	<.001
e-scooter use (used, not used)					-.141	.003
2. Step	.46	.28	5,433	45.45		<.001
Attitude					.210	<.001
Perceived behavioural control					.335	<.001
Perceived usefulness					.175	<.001
Subjective norms					.087	.025
Facilitating conditions					.075	.047

Discussion

In this study, the behavioural intention and factors affecting the use of e-scooters were examined among young people in Turkey. The principal component analysis results showed that the majority of the distinct factors of the Theory of Planned Behaviour and the Technology Acceptance Model manifested individual constructs. In line with Ajzen's (2020) discussion, the factorial structure revealed that the perceived behavioural control items also included perceived ease of use items. Ajzen (2020) identified control factors as "*skills and abilities; availability or lack of time, money, and other resources; cooperation by other people; and so forth*" (p. 315). Perceived ease of use could be assessed in a similar way to perceived behavioural control, as the items indicate the ease of using an e-scooter in a way that indicates the users' control beliefs over e-scooter use.

Comparing users and non-users, significant differences were found in individual factors (i.e., technology adoption, attitudes, perceived behavioural control, perceived usefulness, and behavioural intention). The study was carried out during the early phase of e-scooter use in Turkey. Therefore, it is not surprising that the e-scooter users had higher technology adoption and more positive attitudes and beliefs towards e-scooters in comparison to non-users. In addition, the greatest difference between users and non-users of electric scooters was observed on perceived behavioural control. Perceived behavioural control was also the strongest predictor of behavioural intention. It can be argued that the lack of perceived behavioural control over e-scooter use, combined with the perception that e-scooters are difficult to use, may be the greatest barrier for current non-users of e-scooters.

On the other hand, e-scooter users and non-users did not differ on technical or social aspects (i.e., subjective norms and facilitating conditions). The reason for this difference could be related to the level of integration of e-scooters into everyday life. Similar to previous studies which failed to find the impact of facilitating conditions (Öztaş Karlı et al., 2022), in this study facilitating conditions showed the weakest significant impact on behaviour prediction, with the lowest average score across our samples. This may indicate that the infrastructure and policy readiness of the transport system for e-scooters might not have been sufficient to make a difference in the users' mode choice yet. Similarly, various studies have reported a lack of suitable infrastructure as being one of the barriers for e-scooter use (Almannaa et al., 2021; Buehler et al., 2021; Rejali et al., 2021; Sanders et al., 2020). Official regulations regarding e-scooters were only published in April 2021 (Resmi Gazete, 2021), which corresponds to the last two months of our data collection. However, the hierarchical regression also showed that both subjective norms and facilitating conditions are important factors in behavioural intention. This could mean that positive subjective norms towards e-scooter use and improved facilitating conditions could play a key role for a certain group of young people to start (or increase) the use of e-scooters.

In line with previous studies (e.g., Javadinasr et al., 2022; Öztaş Karlı et al., 2022; Ratan et al., 2021; Rejali et al., 2021), all predictors were positively related to the behavioural intention to use e-scooters in the near future. The results of this study suggest that the theory of planned behaviour, along with the additional constructs (perceived usefulness and facilitating conditions), are useful in explaining the adoption of electric scooters. The results also imply that individual factors such as attitude, perceived behavioural control, perceived usefulness, and behavioural intention are important in predicting e-scooter adoption. The results of this study can be used to inform policymakers and transport planners about the importance of these factors in increasing e-scooter use among young people.

This study is limited in the following ways: First, the current study did not differentiate the types of e-scooter use/ownership (e.g., owning an e-scooter, using shared transport apps). Future studies may control certain aspects of e-scooter use to make inferences with higher detail, such as e-scooter apps, ease of access, and parking. Second, the sample of the current study is limited to young people between the ages of 18 and 25. The generalisability of the results may be limited to this group of road users, given differences with other age groups in terms of transport activities and other key factors pertaining to road use. Therefore, there is a need for further research with varying age groups to improve the generalisability of the results.

Finally, the findings of this study suggest a number of theoretical and practical implications. Theoretically, in addition to the original theory of planned behaviour constructs (attitude, subjective norms, perceived behavioural control), facilitating conditions and perceived usefulness were examined in explaining behavioural intention to use e-scooters. This perspective has yielded results that indicate the value of examining different constructs originating from different theories together. Practically, it is also believed that the comparison of users and non-users will help researchers, policy makers and industry to understand the current point of view of e-scooter users and non-users. These findings can be used to improve e-scooter infrastructure, safety, policy, and marketing. A study conducted by Buehler et al. (2021) showed a positive increase in non-users' perceptions of various aspects of e-scooters, such as safety and usefulness, following a pilot project. Public perception and subjective norms of e-scooter use could be improved through controlled pilot projects.

Conclusion

Overall, all aspects considered in this study had a unique and positive effect on behavioural intention to use e-scooters. Perceived behavioural control and attitude were the strongest predictors, and facilitating conditions and subjective norms were the weakest predictors. The results of our study explored the existing difference between e-scooter users and non-users and the usefulness of the constructs in understanding young people's acceptance of e-scooters. The results can be used to inform policymakers, city planners and e-scooter companies in developing strategies to increase the safe use of e-scooters.

Author contribution

İbrahim Öztürk: Methodology, Conceptualisation, Formal analysis, Writing – original draft, Nazlı Akay: Methodology, Conceptualisation, Writing – review & editing.

Acknowledgement

All authors have read and approved to the published version of the manuscript. The authors have no conflicts of interest to declare and have received no funding for this study. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to the author accepted manuscript version of the manuscript arising from this submission.

Data access statement

The data supporting the results of this study are available on request from the corresponding author (I.O., i.ozturk@leeds.ac.uk).

References

Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50(2), 179-211.

Ajzen, I. (2020). The theory of planned behavior: Frequently asked questions. Human Behavior and Emerging Technologies, 2(4), 314-324.

Almannaa, M. H., Alsahhaf, F. A., Ashqar, H. I., Elhenawy, M., Masoud, M., & Rakotonirainy, A. (2021). Perception analysis of E-scooter riders and non-riders in Riyadh, Saudi Arabia: Survey outputs. Sustainability, 13(2), 863.

Buckley, L., Kaye, S. A., & Pradhan, A. K. (2018). Psychosocial factors associated with intended use of automated vehicles: A simulated driving study. Accident Analysis & Prevention, 115, 202-208.

Buehler, R., Broaddus, A., Sweeney, T., Zhang, W., White, E., & Mollenhauer, M. (2021). Changes in travel behavior, attitudes, and preferences among e-scooter riders and nonriders: first look at results from pre and post e-scooter system launch surveys at Virginia Tech. Transportation Research Record, 2675(9), 335-345.

Chen, C. F., & Chao, W. H. (2011). Habitual or reasoned? Using the theory of planned behavior, technology acceptance model, and habit to examine switching intentions toward public transit. Transportation Research Part F: Traffic Psychology and Behaviour, 14(2), 128-137.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340.

Javadinasr, M., Asgharpour, S., Rahimi, E., Choobchian, P., Mohammadian, A. K., & Auld, J. (2022). Eliciting attitudinal factors affecting the continuance use of E-scooters: An empirical study in Chicago. Transportation Research Part F: Traffic Psychology and Behaviour, 87, 87-101.

Madigan, R., Louw, T., Wilbrink, M., Schieben, A., & Merat, N. (2017). What influences the decision to use automated public transport? Using UTAUT to understand public acceptance of automated road transport systems. Transportation Research Part F: Traffic Psychology and Behaviour, 50, 55-64.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. Behavior Research Methods, Instrumentation, and Computers, 32, 396-402.

Öztaş Karlı, R. G., Karlı, H., & Çelikyay, H. S. (2022). Investigating the acceptance of shared e-scooters: empirical evidence from Turkey. Case Studies on Transport Policy, 10(2), 1058-1068.

Petzoldt, T., Ringhand, M., Anke, J., & Schekatz, N. (2021). Do German (Non) Users of E-Scooters Know the Rules (and Do They Agree with Them)?. In HCI in Mobility, Transport, and Automotive Systems: Third International Conference, MobiTAS 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings (pp. 425-435). Cham: Springer International Publishing.

Ratan, R., Earle, K., Rosenthal, S., Chen, V. H. H., Gambino, A., Goggin, G., ... & Lee, K. M. (2021). The (digital) medium of mobility is the message: Examining the influence of e-scooter mobile app perceptions on e-scooter use intent. Computers in Human Behavior Reports, 3, 100076.

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. Psychological Assessment, 12(3), 287.

Rejali, S., Aghabayk, K., Mohammadi, A., & Shiwakoti, N. (2021). Assessing a priori acceptance of shared dockless e-scooters in Iran. Transportation Research Part D: Transport and Environment, 100, 103042.

Resmi Gazete (2021). Elektrikli skuter yönetmeliği. Retrieved from: https://www.resmigazete.gov.tr/eskiler/2021/04/20210414-3.htm

Sanders, R. L., Branion-Calles, M., & Nelson, T. A. (2020). To scoot or not to scoot: Findings from a recent survey about the benefits and barriers of using E-scooters for riders and non-riders. Transportation Research Part A: Policy and Practice, 139, 217-227.

Stevens, J. P. (2009). Applied multivariate statistics for the social sciences. London: Routledge.

Human factors approach to phlebotomy service review

Sharon Beza[1], Lauren Morgan[2], Andrea Granger[3], Joe McCloud [4] & Peter Jeffries[5]

[1]Shrewsbury and Telford Hospitals Trust, [2]Shrewsbury and Telford Hospitals Trust & Morgan Human Systems, [3]Shrewsbury and Telford Hospitals Trust

SUMMARY

In acute hospital care, sampling a patient's blood is frequently used to help guide diagnosis, or to understand a patient's response to treatment. This means many patients will have their blood taken multiple times during an inpatient stay. The work of phlebotomists has been studied before, and acknowledgements made to how they adjust their practice to balance patient safety in the context of fluctuating demands and challenging work environments and equipment (Pickup et al., 2017). A human factors approach was used to analyse the in-patient phlebotomy service within a local National Health Service (NHS) Trust. Multiple systems related issues particularly at organisational level were identified. Recommendations were made on how to improve the safety and reliability of the process.

KEYWORDS

Phlebotomy, safety, HTA, FMEA, SEIPS

Introduction

Phlebotomy, also known as venous blood sampling, is one of the most common invasive clinical procedures. It is an essential tool in diagnosis and treatment of patients. The risk of harm from testing the wrong patient's blood, due to inaccuracies in sample labelling or patient identification is significant, and sometimes results in patient death. A wrong blood in tube (WBIT) incident will influence the likelihood that a patient efficiently and safely receives the required intervention e.g., the transfusion of the correct blood component (Bolton-Maggs et al., 2013). International evidence cited for WBIT incidents is between 1 in every 1,500 – 3,000 of blood samples taken (Cottrell et al., 2013). Other failures in the process have implications for both the patient and the organisation, including delays in treatment and improper utilization of expensive resources (Bolton-Maggs et al., 2015). In view of this and following identification of several problems by phlebotomy staff working on the in-patient service, a decision was made to conduct a human factors or systems-based review of the in-patient phlebotomy service. The aim was to identify any systems related problems that could be addressed to improve process reliability and patient safety, as well as improving the experience of phlebotomy staff, thereby improving the service, as a whole.

Method

The work was conducted over two hospital sites. A multi-methods approach was used to collect data and analyse it. Phlebotomists working at the two hospital sites were observed whilst conducting their daily in-patient /ward phlebotomy rounds in various hospital wards over several days. Observations were conducted for the duration of the phlebotomists shift on each occasion i.e., from 08:00 hours -12:00hours. Phlebotomy outpatient services are also provided at both hospital sites, but these were not included in the study. This was followed by observing ward-based doctors and emergency department staff, over several days. All completion and submission of blood test requests by doctors, as well as all bleeding of patients conducted whilst the observer was on the ward or in the emergency department were included in the study. Observations were combined with semi-structured interviews to further explore in any issues identified. The Systems Engineering Initiative for Patient Safety Framework is a systems analysis tool widely used in healthcare which explores the work system, processes, and outcomes. Findings from the observations and semi-structured interviews were classified according to the categories of the SEIPs work system classification.

Hierarchical task analysis is a popular task analysis tool which 'describes the task under analysis in terms of a hierarchy of goals, sub-goals, operations, and plans. It can be used with other human factors analysis tools in varied ways including design and evaluation, workload assessment and error prediction and analysis. In this study, hierarchical task analysis was used to map out the following key tasks: registering a blood test request (including printing the request form), bleeding patients and processing blood samples in the pathology laboratory.

Failure modes effect analysis (FMEA) is a proactive risk management tool used to identify prospective failures within processes or products, before they occur, and which focuses on system design. It was used to determine failures that could occur in the tasks listed above and the effects of these failures. The failure modes were prioritised using a risk rating matrix. This was based on the frequency of the failure mode and the severity of the effects of the failure, multiplied together to generate a risk priority number (RPN). In addition, the failure modes were ranked in terms of the perceived 'ease of fix'. The RPN and 'ease of fix' were used to determine prioritisation of addressing failure modes.

Findings

Findings were analysed using the Systems Engineering Initiative for Patient Safety (SEIPS) model and fell mainly within the tools and technology and organisational factors work system categories. The following solutions have been identified:

1. Labelling of printers so that request forms are placed in the correct orientation for labels to print on correct side.
2. Extend size of label templates so that all required patient details are captured on the label sticker.
3. Use of the same standard and quality wrist ID bands in all clinical areas.
4. Handover ward list of patients bled and not bled to ward staff by phlebotomists at end of each session.
5. Exploring ways in which all relevant clinical staff can have access to the electronic blood request system.

Conclusion

Human Factors tools, used to analyse a phlebotomy service can identify a significant number of recommendations for risk reduction that the service was unaware of before study. The improvements in phlebotomy service, will impact patient safety throughout the whole hospital.

References

Bolton-Maggs, PHB, Poles D, Watt A et al on behalf of the Serious Hazards of Transfusion (SHOT) Steering Group. The 2013 Annual SHOT Report 2014

Bolton-Maggs PHB, Wood EM, Wiersum-Osselton JC (2015). Wrong blood in tube - potential for serious outcomes: can it be prevented? *Br J Haematol*, 168, 3–13.

Carayon et al (2006). Work system design for patient safety: the SEIPS model. *Qual. Saf Health Care* 15 Suppl1: i50-8.

Cottrell S, Watson D, Eyre TA et al. Interventions to reduce wrong blood in tube errors in transfusion: a systematic review. *Transfus Med Rev* 2013, 27, 197–205

Haroun, A., Al-Ruzzieh, M, A., Hussien, N., et al (2021) Using Failure Mode and Effects Analysis in improving Nursing blood sampling at an international specialized cancer centre. *Asian Pacific Journal of Cancer Prevention, 22, pp.1247-1254.*

Kaufman, R. M., Dinh, A., Cohn, C.S., et al (2019) Electronic patient identification for sample labelling reduces wrong blood in the tube errors. *Transfusion*, 59, pp 972-980

Pickup, L., Atkinson, S., Hollnagel, E., Bowie, P., Gray, S., Rawlinson, S. and Forrester, K., 2017. Blood sampling-Two sides to the story. *Applied ergonomics*, *59*, pp.234-242.

Stanton, N. (2006) Hierarchical Analysis: Developments, applications and extensions. *Applied Ergonomics* 37.pp55-79

Villafranca, JJA., Sanchez,AG., Guindo, MN., et al (2014): Use failure mode and effects analysis to improve the safety of parenteral nutrition. *Am J Health System Pharm*, 71, pp1210-18.

The systemic causes of medication problems for hospitalised children

Adam Sutherland[1 2 3], Suzanne Grant[4], Stephen Tomlin[5]; Denham L Phipps[1 2], & Darren M Ashcroft[1 2]

[1]NIHR Greater Manchester Patient Safety Translational Research Centre; [2]Division of Pharmacy & Optometry, Faculty of Biology Medicine & Health, University of Manchester; [3]Pharmacy Department, Royal Manchester Children's Hospital, Manchester University NHS Foundation Trust; [4]School of Medicine, University of Dundee; [5]Children's Medicines Research Centre, Great Ormond Street Hospital NHS Foundation Trust

SUMMARY

Medication processes are chaotic and complex, and assumed to be undertaken by specific professionals in isolation from other healthcare tasks. However tasks are delivered simultaneously and adaptively because of the complexity of healthcare provision. This study aimed to explore the systemic contributory factors to medication related problems in children's wards using multiple qualitative methods (230 hours participant observation and 19 semi-structured interviews). There is insufficient resource available to undertake all the processes to ensure safety; decisions about medicines were made with reference to immediate problems only; parents were relied on to administer medicines to children, and; there was widespread non-compliance with interventions to improve safety because they conflicted with day-to-day work.

KEYWORDS

Medicines Safety; Ethnography; Children and Young People

Introduction

Medication causes up to 25% of avoidable healthcare harm (Panagioti et al., 2019). Simple interventions have been shown to have limited impact.(Maaskant et al., 2015). Medication safety systems have been described as complex and chaotic.(Hawkins & Morse, 2022) A work domain analysis of medication systems in this setting described a complex interconnected system. (Sutherland et al., 2022.). This study set out to explore the contributory factors that emerge and interact in the system.

Method

A multicentre ethnographic study was conducted across three acute paediatric wards in the north of England. 230hrs of non-participant ethnographic observation and 19 semi-structured interviews with families, pharmacists, medical and nursing staff were carried out. Data was managed and coded using NVIVO v12 (QSR International) and analysed using an inductive thematic analytical approach. Coding and interpretation were undertaken by all members of the research team and agreed by consensus.

Results

Three systemic themes were identified that contributed to medication related problems – logistical issues, cognitive and decision-making processes, and situational and physical environment.

Logistical and resource limitations.

There were multiple processes to ensure safe medicines management, but there were insufficient people or time available to deliver them. Organisations attempted to expand the staff pool for some medication processes which worsened deficiencies in others. Families were relied on by staff to administer medicines to ensure medicines were administered on time but were not formally part of the system. Information technology equipment was often in short supply or inoperable. Prescribing decisions were made on ward rounds by medical teams, and nursing staff were the final arbiters of medication administration. Pharmacy services were reactive; focussing on verification of prescriptions for dispensing. Prescribing intentions were seldom documented leading to challenge and clarification by nursing and pharmacy staff.

Cognitive and Decision-making Processes

Medication processes were considered as separate tasks including prescribing, dispensing and administration. However, none of these tasks were carried out in isolation or independently. Interventions introduced to improve medicines safety (barcode medication administration systems, independent checking) did not integrate with other existing workflows or were inoperable within the circumstances of use. All staff had to make complex clinical decisions with every medication order administered and nursing staff perceived that they were accountable for all medication problems because they administered them.

Physical and situational environment

Medication work was perceived as task focussed with all staff grouping medicines tasks with other duties to make the best use of their time. Medication assessment by prescribers was only in the context of the immediate problem and other issues were picked up by parents, nursing or pharmacy staff. Parents also administered medicines without prescription if deemed in the best interests of their child. Medication spaces were often inadequate for needs – preparation rooms lacked computer access and many medicines were manipulated at nurse's stations. Interruptions were constant and interventions to reduce them unused. Economics drove organisational medicine choices with no consideration of their acceptability to children.

Discussion

This study supports the insights into clinical workload and turbulence in other healthcare settings. (Jennings et al., 2022). Organisations viewed medicines safety as an isolated concept, yet the work was intrinsically linked to the wider system and frontline care. Interventions are implemented within this isolated frame. Furthermore, medicines were viewed only in the context of the acute problem being treated, other potential medication related problems were uncovered by chance. There was also a noted absence of teamwork where medications were being prescribed and administered, resulting in healthcare staff interrupting their work to clarify choices and intent.

References

Hawkins, S. F. & Morse, J. M. (2022). Untenable Expectations : Nurses ' Work in the Context of Medication Administration , Error , and the Organization. *Global Qualitative Nursing Research*, *9*, 1–17.

Jennings, B. M., Baernholdt, M. & Hopkinson, S. G. (2022). Exploring the turbulent nature of

nurses' workflow. *Nursing Outlook*, *70*(3), 440–450.
Maaskant, J. M., Vermeulen, H., Apampa, B., Fernando, B., Ghaleb, M. A., Neubert, A., Thayyil, S. & Soe, A. (2015). Interventions for reducing medication errors in children in hospital. *The Cochrane Database of Systematic Reviews*, *3*, CD006208.
Panagioti, M., Khan, K., Keers, R. N., Abuzour, A., Phipps, D., Kontopantelis, E., Bower, P., Campbell, S., Haneef, R., Avery, A. J. & Ashcroft, D. M. (2019). Prevalence, severity, and nature of preventable patient harm across medical care settings: systematic review and meta-analysis. *BMJ*, *366*, l4185.
Sutherland, A., Phipps, D. L., Gill, A., Wolffsohn, K., Morris, S. & Ashcroft, D. M. (2022). A work domain analysis of medicines management for hospitalised children. In N. Balfe & D. Golightly (Eds.), *Ergonomics & Human Factors*. CIEHF.

Human Factors Guidance for Robotic and Autonomous Systems (RAS)

Claire Hillyer[1], Hannah State-Davey[1], Nicole Hooker[1], Richard Farry[1], Russell Bond[2], James Campbell[2], Phillip Morgan[3], Dylan Jones[3], Juan D. Hernández Vega[3] & Philip Butler[3]

[1]QinetiQ, UK, [2]BMT, UK, [3]Cardiff University, UK

SUMMARY

This paper outlines recent (2021/2022) work to produce Human Factors (HF) guidance to support the design, development, evaluation, and acquisition of Robotic and Autonomous Systems.

KEYWORDS

Robotic, autonomous, system, human factors, human-centred design, artificial intelligence, system

Background and Context

Robotic and Autonomous Systems (RAS)[1] will play an increasingly important role in Defence Capability and they are likely to have a fundamental impact on the way in which future military activities, across the spectrum, are conducted (e.g. combat, humanitarian relief, cyber operations and operational support functions). This will have consequences on, for example: how people interact with these systems; the skills required to acquire, operate and maintain them; and the number, organisation, and location of these personnel. Understanding how to optimise both the human and technological components of such systems is critical.

While significant systematic research has been conducted into RAS technologies, relatively little *human science* research has been conducted outside of generic work on Human-Computer Interaction (HCI) or on topics such as the safety of Remotely Piloted Vehicles (RPVs). There is still a lot to learn about how humans might interact with complex RAS and how these technologies might be, optimally, integrated with the human component, both at the individual and at the collective level (such as might be reflected in Human-Machine Teaming, HMT[2]).

Development of RAS is progressing at pace and timely guidance, covering the Human Factors (HF) considerations associated with these systems, is essential to support Defence to design, field and operate RAS and build appropriate trust in these systems.

Exploration Guide – 'Human Factors Considerations for the Development and Testing of Robotic and Autonomous Systems'

In 2021/2022, a study was conducted to produce an Exploration Guide to raise awareness of the HF opportunities and challenges associated with the use of RAS by Defence and to provide advice and guidance on how to respond to these. Key considerations include: Situational Awareness (SA) and

[1] RAS is an accepted term used by academia and the science and technology community to highlight the physical (robotic) and/or cognitive (autonomous) aspects of a system (or platform).

[2] HMT is, essentially, a relationship – one made up of at least three equally important elements: the human, the machine and the interactions and interdependencies between them.

workload; system trust and reliability; decision transparency and explainability (i.e. what is the logic, process, factors or reasoning upon which an Artificial Intelligence (AI)-enabled system's actions or recommendations are based); physical design considerations; and HMT and communication. It draws on the outputs of a comprehensive review of over 300 publications including empirical research, published best practice, and standards and guidelines examining the design and operation of systems (within both military and non-military domains). Initially, a set of 60 search terms was developed and 1014 papers identified for review (via Google Scholar). The team conducted a relevance scoring activity to filter the papers down to a core set of 272 papers and this set was supplemented with additional information sources (including the Defence Standard (Def Stan) 00-251 'Human Factors Integration for Defence Systems' Technical Guides). The final set of papers was subject to a more detailed review and key themes were identified, for inclusion in the final Guide.

Figure 1 shows example excerpts from the Exploration Guide.

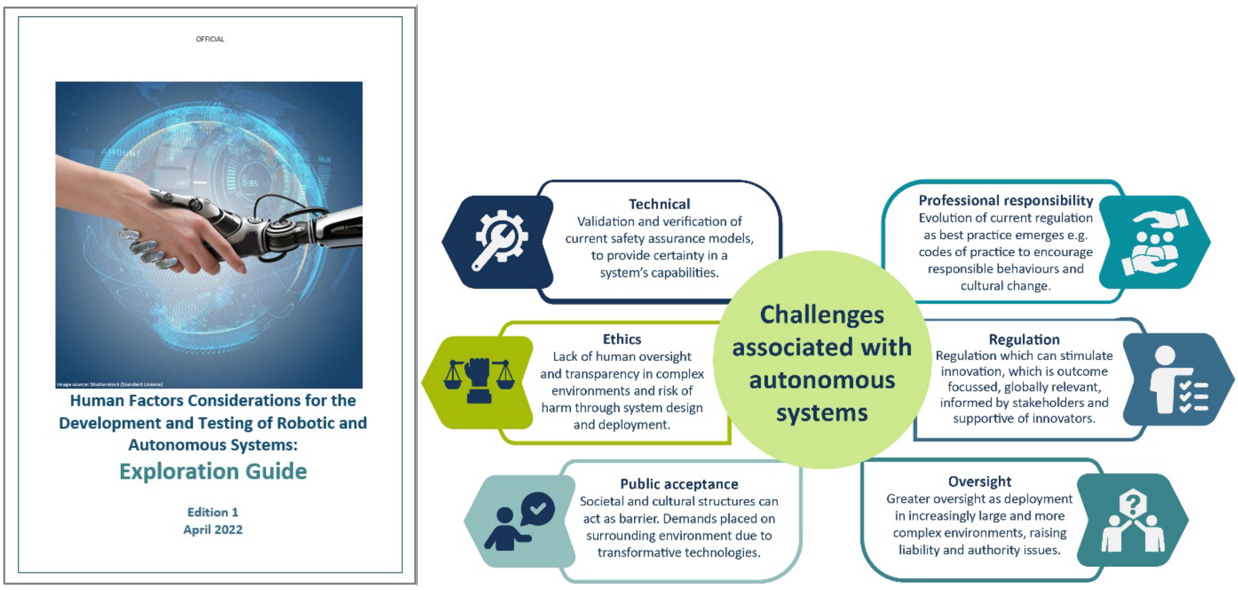


Figure 1: Exploitation Guide excerpts – front page and an example visualisation

RAS present a unique set of HF considerations over those that apply to other systems. The interaction between a human operator and a RAS is different from typical Human-Computer Interaction (HCI) as they may employ AI, involve complex dynamic control systems, exhibit high levels of autonomy and operate in changing real-world environments. This Exploration Guide provides high-level guidance to address many of these challenges and opportunities.

Who is the intended audience?

This Guide has been designed for use, primarily, by HF practitioners involved in developing, evaluating, acquiring and/or commissioning RAS. However, outside of this primary audience, it also provides useful contextual and awareness material for systems and software designers and engineers less familiar with basic HF approaches.

Exploitation

Work is in progress to exploit this output by uploading it onto the Knowledge in Defence Human Factors Integration Management System (HuFIMS) to sit alongside other Human Factors Technical Guides.

Digital Simulation Modelling providing a platform for ETCS Driveability Assessments

John Gunnell

Arup, UK on behalf of Network Rail UK

ABSTRACT

The development of a true-to-life, cutting-edge Digital Simulation Model of a train drivers experience operating with a future European Train Control System (ETCS) system has provided a platform for robust Human Factors analysis and has proven to be a success with the driver stakeholders.

KEYWORDS

Digital, Simulation, Model, Visualisation, Rail, Driveability, ETCS, Look-and-Feel, Innovation

Introduction

The digitalisation of the UK's Transpennine railway will see a transformational shift from conventional manual operations of train signalling to a semi-automated operation through introducing enhanced technology offering system protection control. ETCS is an in-cab signalling system allowing trains to run closer together, safely and to travel at their optimal speeds.

A driveability assessment has been carried out during the concept design phase of the Transpennine Route Upgrade (TRU) ETCS project to assess the route from the train driver's perspective in the new operating environment and conditions.

The cost and time for implementing digital signalling is under heavy scrutiny since Network Rail initiated their 'Target 190' industry-wide program to provide the capability to enable safe, affordable, and deliverable signalling to meet the future demands of the railway, (Network Rail, 2023). Digital simulation models can have multiple benefits to the rail industry. TRU project identified set of opportunities to harness the technology in conducting the driveability assessments, as presented in Figure 1.

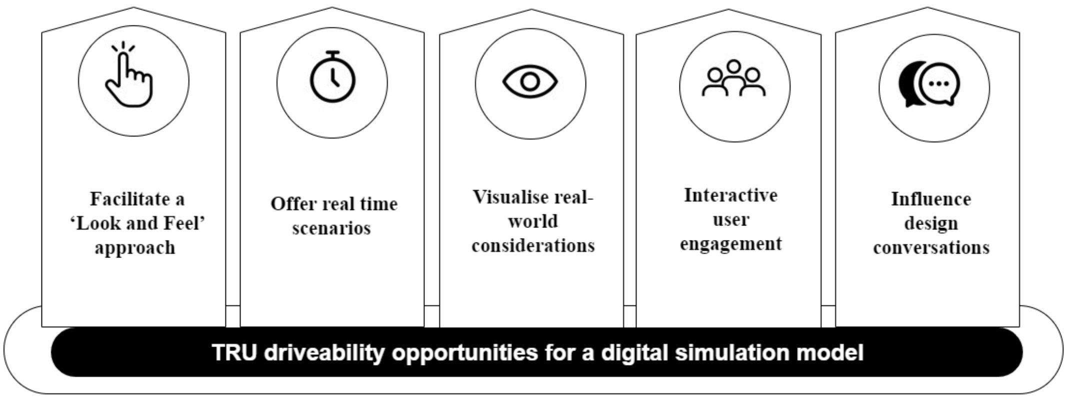


Figure 1: TRU driveability opportunities for a digital simulation model

The digital simulation model, Rail Signalling Visualisation Tool (RSVT), was developed by Arup UK on behalf of Network Rail to replicate the future state and present a representative 'look and feel' of the train driver's in-cab view. The RSVT allowed for virtual design reviews for the TRU in-cab signalling project. As a cornerstone of the driveability assessment approach the RSVT provided a multitude of benefits to the project analysis in operations, human factors, safety, and performance.

A traditional signalling design review process based on drawings is sub-optimal, significant signalling system experience is required, otherwise it can be very difficult for engineers and operators to interpret. The RSVT, a next generation digital simulation model of the rail environment provided a high-quality resolution rendering of the real world, whist augmenting the future state infrastructure using BIM (Building Information Model) digital models. The use of an off-the-shelf simulation package meant that the train movements could be calculated to represent the operational line speeds and motions over various operating scenarios.

The generation of RSVT, provided a valuable signalling concept design reference tool for stakeholders integration throughout the project user engagements. During the early project design utilising the innovative tool enabling understanding and acceptance without significant experience of operating cab signalling systems.

RSVT modelled scenario recordings benefitted the human factors task analysis and human error analysis methods applied to the project to best identify the preconditions in the concept design that give rise to errors before they occur. A detailed systematic review of the driver to rail system interactions could be conducted to understand the ETCS demands and potential task conflicts across the board harnessing the flexible replay and time-based specific functions particular to the RSVT.

Method

The TRU project driveability study set out to review the risks introduced by the conceptual ETCS safety system and determine how they should be appropriately managed through the design functions. It involved the development of the RSVT digital simulation model, a series of stakeholder engagement workshops, and an evaluation of the proposed future driving task with ETCS in operation.

RSVT Development

The RSVT model aimed to provide the complete future state route in a virtual design environment with the visual perspective of the in-cab driver. The tool set out to offer the ability to map the driver Driver Machine Interface (DMI) screen to the external environment. A mature visualisation tool was possible by inputs by connecting the environmental BIM modelling, high quality 3D rendering, and route specific train simulator technology.

Design development of the RSVT digital simulation model required an interactive and iterative process. It was imperative that the detail of the future route and driver DMI was accurately portrayed and any discrepancies with the proposed scheme plan concept could be eliminated. The tool design followed a review process which ensured it captured the proposed signalling scheme plan, existing route information, and future operating conditions to ensure visualisation validity. The operating condition are part of the ETCS reference design requirements specification, (European Railway Agency, 2007).

Driveability Workshops

Stakeholder driveability workshops were conducted to engage with train operators and experienced drivers. A common systematic approach was implemented to all workshops across the impacted areas. The future state train driving operating scenarios were followed and discussed in terms of the hazard precursors, (Rail Industry Standard, 2018). Documentation of any potential impacts identified from the workshop engagement were captured considering the following:

- Signalling information provided
- Display of signalling information
- Positioning of information in relation to driver's field of vision
- Time available to the train driver to comply with operating requirements.

The workshops provided a platform the discuss the current risk mitigations and explore potential mitigations with driver representatives offering a perspective of the effectiveness of the proposed considerations. The RSVT offered the opportunity to view the environment in real-time, pause and play back during key discussion points. It offered a richer level of engagement with the end user group in comparison with the traditional approach where signalling diagrams are reviewed.

The tool allowed the project to model the transitions between ETCS and conventional signalling, the visibility of trackside signage, and to review the route environment to identify any conflicts.

The driveability workshop presented a set of human factors (HF) issues based on simulation scenarios to capture in the project HF issues log. This was to further assess the risks, assumptions, issues, and dependencies identified from the discussion with stakeholders.

Evaluation of proposed future driving tasks

A literature review of the driveability analysis of ETCS, (Rosberg et al, 2021), and the transition to/from ETCS operations, (Rail Safety and Standards Board, 2016), provided examples of structured approaches to evaluate the future driving task considering the technology proposed to be introduced. RSSB produced a task analysis and a list of plausible driver errors for transitions across different signalling systems.

For TRU, a tabular task analysis captured the current driving duties in the impacted section and the driving duties in the proposed future design for normal and degraded operations, including the level transitions. Each task was decomposed to its lowest level of action, based on the information available to the project. These individual action were considered in terms of the sensory, cognitive, or psychomotor activity, required by the driver to achieve the task.

All plausible human errors from the task analysis were systematically reviewed to determine the measures required to mitigate the error producing conditions and minimise the consequences. Having the ability to share physical evidence from the RSVT results, dissemination of the identified risks with the designers and drivers was a more transparent, simpler validation process.

Results

Driveability Workshops

A summary of the driveability workshop findings particularly benefiting from the RSVT simulation model can be found in Table 1;

Table 1: Summary of driveability workshops considerations utilising the RSVT

Consideration Type	Consideration Description
Signalling information provided	Timing of the ETCS transition announcement indicator conflicts with existing Automated Warning System (AWS) acknowledgement information. ETCS indications expects not interfere with existing demand, shown in Figure 2.

Display of signalling information	DMI-monitored speeds should be aligned to the available trackside speed boards. The review should consider if any change to the speed board is required due to the identified potential overspeed.
	A release speed may be mistaken for an Movement Authority extension.
Positioning of information in relation to driver's field of vision	First signal at the transition point from ETCS to non-ETCS not visible due to track infrastructure and geography.
	Sighting of degraded Stop Marker Boards should be unrestricted to ensure drivers do not overspeed the maximum permitted speed under degraded working, as shown in Figure 3.
	Neutral Section additional driver warning needed. Driver is instructed to follow the instructions from the DMI, lineside signage would reconfirm the upcoming neutral section.
Time available to the train driver to comply with operating requirements	Consistency of the upcoming ETCS transition indication announcement timings with other ETCS applications across the network.

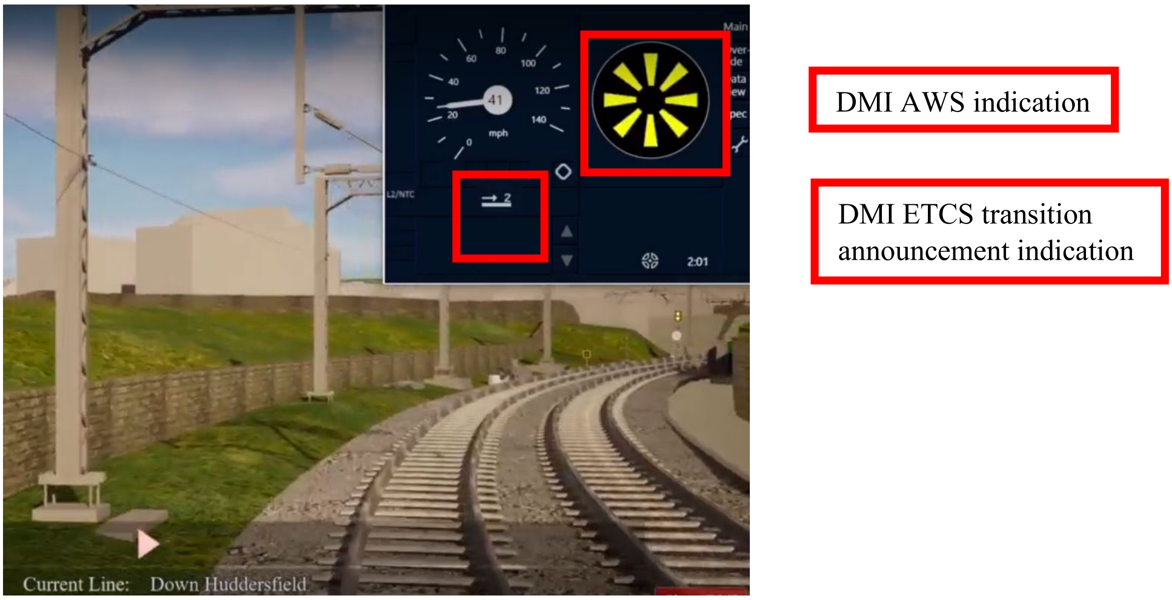


Figure 2: DMI ETCS transition announcement indicator and AWS task conflict

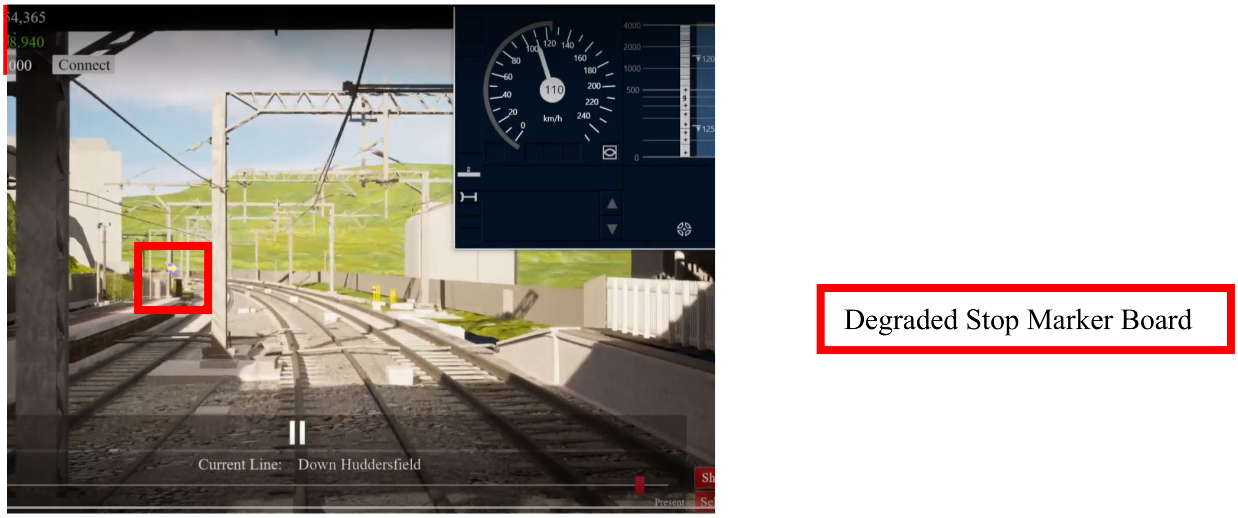


Figure 3: Degraded Stop Marker Board restricted sighting

Evaluation of proposed future driving tasks

The RSVT was used to review the future normal and degraded driving tasks whilst considering the existing driving functions to remain. Further physical and cognitive task conflicts were identified in addition to those captured in through the stakeholder engagement driveability workshops. Figure 4 and Figure 5 show two key considerations identified in the task analysis through the RSVT include;

Figure 4: DMI speed change indication distraction of station stopping duties

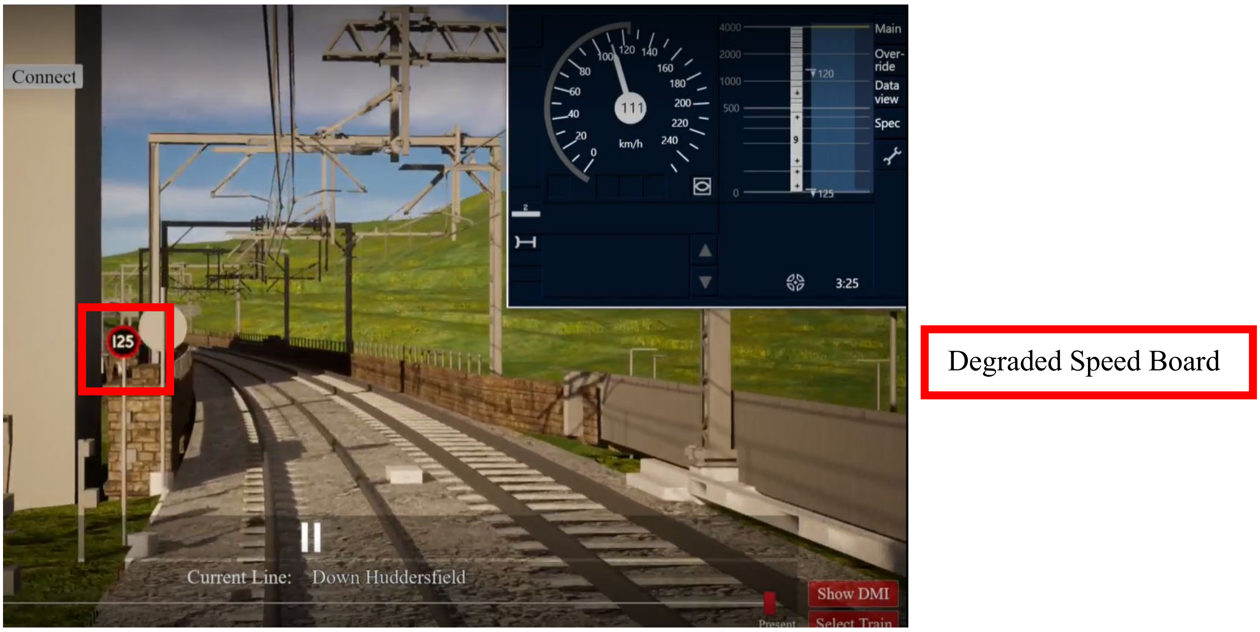


Figure 5: Presentation confusion of Degraded Speed Board

The human error analysis identified a set of plausible errors from the systematic review of the proposed future state driving task in the RSVT. A summary of key task conflicts or design limitations to potentially effect operations detected from the RSVT review are listed in Table 2. The table outlines the task impacted; the potential plausible errors; the TRU context specific error producing conditions; any safety or performance consequences; recovery opportunities available to the driver; and recommended project design and organisational measures to mitigate the error risks.

Task	Plausible Error	Error producing conditions	Consequence(s)	Recovery opportunities	Measures
Normal Operation					
Hear/sight ETCS speed change information on DMI	Information not obtained - Missed indication or not seen Check omitted- Driver fails to process the indication Operation mistimed	Conflict between the speed change indication presented at the same time as preparing for station duties. At location x speed change ETCS information estimated 15 seconds prior to speed change lasting 20 seconds. The indications expect to be present until approx. 10 seconds after location x. Drivers attention may be focused on approaching station duties and fail to process the information demands and operate the train faster / slower than the new limit permits.	Performance: Driver not prepared/anticipating speed change, and associated control required. Performance: Sudden braking from the system could result in harsh jolting of the train and impacting the passenger experience.	Both audible (beep beep) and DMI indicator presented. In some locations Driver's station duties expect to involve braking aligning with the intentions of the speed change indications. Once the speed limit has changed the hook will constantly show the new limit.	Suitable spacing of speed change indications away from approaching station. Driver competency assessment. Driver route knowledge.
Visualise approaching speed change in planning window	Information not obtained - missed speed change indication Operation mistimed	Internal focus on DMI speed indicator different to existing conventional speed change task. Speed change positioning prior to location x conflicts with Driver's demand to brake for the station. Speed increase missed by driver due to station duties.	Safety: Driver misses the prompt to decrease speed resulting in sudden braking from driver risking passenger safety. Performance: Service brake intervention control input stopping train and slowing the service.	Service brake intervention will prevent the driver from overspeeding. Station duties will require the drive to reduce the speed. Route knowledge of the speed change and station positioning not expected to change unless station proximity has changed.	Suitable spacing of speed change indicator from stations . Driver competency assessment. Route Knowledge.
Degraded Operation					
Sight degraded Speed Board (kph)	Check omitted Wrong information obtained	No speed board signage provided for the degraded trains after speed reduction in identified locations. Driver continues to operate at lower speed that permitted. Driver confusion as to where the permitted speed changes and increases the speed too early.	Performance: Driver continues to operate at a lower speed instructed. Delay to operations impacting network performance. Safety: Train exceeds the permitted speed risking train derailment.	Written order instructs driver to operate at a permitted speed unless instructed otherwise. Driver route knowledge expected to return to permitted speed at certain locations. Communication to signaller available to confirm changes in linespeed if Driver is unsure.	Signage design to be considered and positioned to signal sighting guidelines. Clear written order communication of the permitted speed.
Sight Km post	Wrong information obtained Wrong information communicated	Driver confusion between km posts and mile post coexisting trackside in ETCS areas.	Safety: Drive misinterprets the nearest signage when communicating the train location to the Signaller in degraded/emergency scenarios. Impact on signalling movements of train into potentially occupied track.	Design differences between the km posts introduced and the existing mile post. Yellow background and Km text to be included.	Signaller and Driver to confirm the train location unit. Consider the removal on Mile Posts in ETCS areas to avoid confusion. Route Knowledge.

Table 2: Summary of plausible human errors identified from the RSVT review process

Discussion

The driveability workshops and driver task evaluation design considerations prompted a series of risks mitigation design development hazard analysis reviews. A key ETCS design aspect is the positioning of the level transition from ETCS to non-ETCS train operation modes is required to offer the driver visibility of the first signal. The simulation model offered an opportunity to investigate how fit for purpose the design is from a drivers perspective. The RSVT demonstrated limitations of the concept design in a specific TRU context, due to the trackside infrastructure and track geography. The visualised early detection has led to the project reviewing the risk to driver and explore potential improvement opportunities in the advance project stages.

The results from driveability workshops identified the difficulty to sight degraded Stop Marker Boards in specific locations where the track curvature restricted the approaching driver's line of sight. Traditional concept design driveability reviews would likely fail to capture the risk where representative visualisation of the section would be absent. The project can utilise the RSVT during the development with the ability to trial repositioned boards in the virtual world before finalising locations. A similar approach is available with other physical ETCS assets introduced such as Cab Boards, degraded Speed Boards, Km Posts.

The participating stakeholders feedback was of an enhanced sighting experience through the accurate representation of the real-life rail environment. By making the simulations 'Look and Feel' so realistic it harnessed the personal ability and knowledge from the drivers, improving the collaboration and input to design through numerous context specific risk consideration generated.

The model provided a mechanism to accurately breakdown the proposed driver workload and demand in the design of future state operating conditions and produce a set of error vulnerabilities identified for driving in ETCS. By enhancing technical understanding of the functional changes for the train operators, the model uncovered design considerations of split attention and overloading around level transitions. The analysis has provided a mechanism to development the TRU specific signalling design and outline specific system design requirements.

RSVT will play an important role in the development of digital twins. There is an expanding industry need for digital twins for the railway and this is a trend that will only grow. RSVT allows change development, testing and visualisation within the digital twin first. This means solutions can be tested quickly and effectively, and thereby reducing the cost and time for deployment on the real railway.

Study Limitations

The application of the RSVT digital simulation model on TRU has three notable limitations. First, the RSVT driver interface information provided did not contain the full current and future state in-cab messaging, alarms and communications requiring the drivers demand and attention during the driving task. Secondly, the level of simulation fidelity varied across the TRU scheme impacting the representation clarity of the trackside environment in sections. The quality of the RSVT is reliant on consistent and accurate data inputs. Third and finally, the observer expectancy effect. The perceived expectations of the future operation can influence the people observing the set RSVT simulated scenarios. False positives portrayed in the scenarios could have influenced the stakeholders expectations and may have led to an unconscious bias.

Conclusion

The driveability assessment approach on the TRU project was enhanced by utilising the digital simulation model RSVT. The resulting design development considerations identified through stakeholder engagement and human factors analysis methods were enriched by the capabilities of

RSVT. It is recommended to apply such methods of technology to future signalling design reviews, both to detect valuable design risks in the early project stage and to improve the stakeholder engagement experience.

References

European Railway Agency (2007). ERTMS/ETCS Functional Requirements Specification ERA/ERTMS/003204. 5.

Network Rail (2023). Target 190plus: Sustainable signalling renewals [online] https://www.networkrail.co.uk/industry-and-commercial/research-development-and-technology/research-and-development-programme/target-190plus-sustainable-signalling-renewals/

Rail Industry Standard (2018). Lineside Signalling Layout Driveability Assessment Requirements RIS-0713-CCS. 1.1

Rail Safety and Standards Board (2016). Transition to/from ERTMS operations, Research and development

Rosberg, T, Cavalcanti, T, Thorslund, B, Prytz, E, Moertl, P. (2021) Driveability analysis of the European rail transport management system (ERTMS): A systematic literature review Journal of Rail Transport Planning & Management 18

Defining Roles of the Remote Operator in Autonomous Vehicles

Hannah Parr[1], Catherine Harvey[1], Gary Burnett[1]

[1]University of Nottingham

SUMMARY

This paper proposes four distinct roles of a remote operator in supporting future SAE Level 4 and 5 autonomous vehicles (AVs). These are Remote Monitoring, Remote Assistance, Remote Management and Remote Driving. A set of scenarios were created based on academic literature and industry reports to represent an extensive range of interactions which might occur between Remote Operators, AVs and other human agents. Operator sequence diagrams were created to represent the task in each scenario and analysed to identify the involvement of the different Remote Operator roles. This is used to draw conclusions on aspects of work for future Remote Operators and as a starting point for further investigation into information requirements and workstation design.

KEYWORDS

Remote Operation, Autonomous Vehicles, Operator Sequence Diagrams

'Levels' of Remote Operation

Although remote operation itself is not a new concept, new challenges are posed by introducing it to on-road vehicles. The focus of this research is autonomous vehicles (AVs) at SAE Level 4 and 5, where the user is not expected/able to monitor or take over control. The aim of this paper is to investigate the different roles of an AV remote operator (RO) through Operator Sequence Diagrams (OSDs) created for a range of possible scenarios. There is currently no widely accepted definition of RO roles. For example, some authors have distinguished between the type of control an operator has over the dynamic driving task (DDT), as either direct or indirect (Kettwich, et al., 2021). Other authors have taken a more task-based approach, distinguishing between remote assistance and remote driving (SAE International , 2021). In this work, a comprehensive literature review was conducted and four 'levels' proposed. These are believed to be representative of all possible RO tasks and take a human-centred approach to describing roles of the RO:

Remote Monitoring (RMo): *Remote observation of AV, user state and environmental factors, supporting the prediction and identification of issues to inform decision making.*

Remote Assistance (RA): *Remote provision of assistance and/or information to the AV user or external agents in close proximity to the AV (e.g. emergency services or vehicle recovery).*

Remote Management (RMa): *Remote provision of instructions to AV to initiate system actions where the AV systems are unable to proceed independently. May also cover fleet management.*

Remote Driving (RD): *Remote control over the dynamic driving task (DDT) of an AV for a limited time period, where RA, RMa and RD are unable to resolve issues of vehicle function.*

These four levels are hierarchal in terms of the extent of influence over the behaviour of the AV. RMo represents the lowest level of influence, through RA, RMa and finally RD at the highest level. This hierarchal structure will be helpful when thinking about the design of work for a RO, where tasks may need to be distributed amongst different operators with different skill sets and/or escalated to higher levels of influence depending on the evolving situation.

Development and Analysis of RO/AV Scenarios and OSDs

To test the proposed levels, a number of scenarios were developed to represent likely future RO/AV interactions. These were based upon academic literature (e.g. Kettwich, et al., 2021) and industry reports (including on-road testing in the US). A final 21 scenarios were identified. Work by Kettwich et al. (2022) compiled a taxonomy of scenarios for remote operation from several context, including observations of public transport control centre staff, videos of road events and interviews with existing AV safety operators. This taxonomy was used to cross reference and validate the scenario list developed as part of this study. From this, 8 scenarios which appeared both in the Kettwich et al. taxonomy and in the list produced as part of this work were selected to represent a range of tasks involving ROs interacting with an AV and human agents in the system, including AV users and external agents.

Previous work by Banks, et al., (2014) successfully demonstrated the use of OSDs to explore the effects of different levels of vehicle automation on system network dynamics. Several cycles of iteration were used to create the OSD's used for analysis, with experts chosen for their experience in AV research and in other Human Factors contexts as a part of this process. For each scenario OSD, roles (RMo, RA, RMa, RD) were assigned to the tasks identified. The frequency with which these roles occurred was then analysed to provide an indication of the relative importance of each. In addition, the OSDs indicated at what points RO roles occur within a scenario and whether they overlap in time.

Findings and Conclusions

The OSDs showed that the roles of RA and RMa occur most frequently in the RO/AV scenarios and the option for RD is required very infrequently (i.e. only when all other roles are unable to resolve the AV control issue). Much research in the sector appears to focus on RD as the main form of remote operation, but the current work suggests that more attention is needed on indirect forms of remote operation (RA and RMa). Remote management appeared most frequently within the analysed OSDs, reflecting the importance of high-level decision-making support for AV behaviour as well as the importance of communication with other system agents and the AV network as a whole. The analysis also showed that RA and RMa often occur simultaneously within a scenario. This has important implications for job design, i.e., to achieve appropriate workloads it may be necessary to have operators taking on different roles and working collaboratively on a single AV scenario, rather than a single RO being responsible for all levels of operation within a geographically-defined area. This also suggests that workstations will need to be designed to support multiple roles and collaborative working. Future work is planned to investigate the information requirements for ROs, particularly at the management and assistance levels, working with Level 4/5 AV services.

References

Banks, V. A., Stanton, N. A. & Harvey, C., 2014. Sub-systems on the road to vehicle automation: Hands and feet free but not 'mind' free driving. Safety Science, Volume 62, pp. 505-514

Kettwich, C., Schrank, A. & Oehl, M., 2021. Teleoperation of Highly Automated Vehicles in Public Transport: user-Centred Design of a Human Machine Interface for Remote-Operation and Its Expert Usability Evaluation. Multimodal Technologies and Interaction, 5(5), p. 26.

Kettwich, C., Schrank, A., Avsar, H., & Oehl, M., 2022. A Helping Human Hand: Relevant Scenarios for the Remote Operation of Highly Automated Vehicles in Public Transport. *Applied Sciences*, *12*(9)

SAE International, 2021. 3016_202104 Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, s.l.: SAE International.

Making the Right Choices: Behavioural Safety for Designers on a Construction Project

Shelley Stiles[1]

[1]Gateway Consultants (HSW) Ltd

SUMMARY

Behavioural safety programmes are widely used across the Construction Industry, largely targeted at influencing behaviours of frontline workers and/or leadership behaviours. However, there is limited application of behavioural safety at the pre-construction (design) phase of a construction project, given the importance of the design community in eliminating and mitigating health and safety risks. This paper details a case study for the application of behavioural safety intervention targeted at the design community for a large infrastructure project.

KEYWORDS

Behavioural Safety, Construction, Design Community

Introduction

In the UK, construction is worth over £100bn and employs over 2.4 million people (Rhodes, 2019). Several studies have identified the cause of accidents to be due to decision made during the planning or design stage of construction projects, with one study in the United States (Behm 2005) identified that 42% of construction site fatalities were linked to design. Designers (both individuals and organisations) are appointed by construction clients to undertake design work which involves preparing or modifying designs for construction projects which may include architects, as well as discipline specialists e.g. structural engineers. Designs include drawings, design details, specifications, bills of quantity and design calculations (CDM 2015). The decisions made by designers fundamentally affect the health and safety of construction work, as well as those operating or maintaining the final asset.

Over the last 15 years construction contractors have been implementing behavioural safety programmes (referred to as BSPs for the remainder of this paper), following on from their reported success in other industries; oil and gas etc. A BSP is a wide term for any activity focused on changing workplace behaviours that cause (or are believed to cause) accidents. Despite considerable research on the contents of BSP's, research to date does not consider the contribution of the design community, and their behaviours, as part of the overall project safety culture. The study objectives are to explore the effectiveness of a behavioural safety training programme specifically for the design community, within a case study.

Case Study: Applying Behavioural Safety within the Design Community

A "Health and Safety Design Risk Management: Making the Right Choices" training programme has been delivered for one large infrastructure construction project valued at £470 million. The total project lasted 9.5 years and at the peak there were 100 designers working on the project. The content of the training programme was developed by Gateway Consultants (HSW) Ltd in line with the project BSP, previously targeted at frontline construction workers. Each training session was delivered by a member of the Gateway Team to members of the design community working on this project from across a number of different design disciplines, and organisations. A blended learning approach was undertaken with a combination of classroom tutorials, e-learning modules and project task activities. The aim of the programme was to focus on good designer behaviours that support the elimination or mitigation of health and safety risks in accordance with the General Principles of Prevention. There were four key behaviours:

- Be alert - lookahead, anticipate and prevent hazard/risk to the health and safety of everyone
- Value user input – involve operators, construction and maintenance teams
- Share information – right information to the right people at the right time
- Keep it real – understand real world application/challenges

A total of 51 individuals from within the design community attended the training sessions scheduled over a 4-month period, spread across 8 cohorts. Feedback was taken from attendees at the end of the training by completing a digital anonymous survey. The feedback showed attendees rated the course as excellent (60%), very good (20%) and good (20%). 97% of attendees stated that the training would be beneficial back at work. With regard the question ‘Which part of the course will be most useful back at work?’ the most common responses include predictive behaviours, human factors, culture of teams, refresh of CDM and managing risks through design. Therefore, it can be concluded that the initial reaction of individuals attending the training session was positive and considered to be of benefit.

Further feedback from the Project Senior Leadership Team have identified improvements following the training interventions with more challenging conversations about health and safety risk management amongst the design community working on the project. Reported examples were built upon the outputs from one of the training activities; where real project design tasks were evaluated against the four key behaviours, and individuals tasked with identifying areas for improvement. This was reported as a particular positive outcome from the training sessions helping to bridge the gap between learning and improving behaviours in the workplace. For example, identifying that meetings were not always an open forum for all attendees to speak up, thus valuable input was missed for a brickwork design scenario. Following the training activity, the individuals established behavioural ground rules for these forums enabling better engagement and greater input from users.

Another tangible improvement from the programme was reported by the Project Director and Engineering Director. Regular designer visits are now being undertaken to the construction site to bridge the gap between design and operational teams which focus on problem solving, shared learning and continue to build on the content provided in the training sessions.

Recommendations

It is concluded that behavioural safety may have a role to play for the design community working on UK construction projects. Further data collection is needed to evaluate whether a BSP intervention can have a significant impact improving safety culture by improving behaviour of designers during the pre-construction stage of a project.

References

Behm, M. (2005) ‘Linking construction fatalities to the design for construction’. Safety Concept

CDM 2015 Available from https://www.legislation.gov.uk/uksi/2015/51/contents/made. Last accessed 23/11/22.

Rhodes, C. (2019) Construction industry: statistics and policy. House of Commons Briefing Paper Number 01432. Available from https://researchbriefings.files.parliament.uk/documents/SN01432/SN01432.pdf Last accessed 23/11/22

Insights into Human Behaviour Hold the Key to the Energy Transition

Kirsty Novis[1] & Thomas A. Norton[1]

[1]Arup Group, UK

SUMMARY

The paper describes how understanding user behaviour and implementing behaviour change initiatives may be the key to changing attitudes to energy consumption and energy transition. The objective of this paper is to explore how four key factors; behaviour centred-design, preventing rebound effects, providing meaningful feedback and, understanding how behaviour scales can make an impact.

KEYWORDS

Sustainability, Energy Consumption, Net Zero, Human Behaviour, Behaviour Change

Introduction

For all the complexity of climate science, the urgent need to transition to renewable energy is widely understood. In 2021, fossil fuel combustion in the global energy sector emitted more than 33 billion tonnes of carbon dioxide – up 5.6% on 2020 (International Energy Agency, 2021), when the world's economy ground to a halt during the pandemic. At COP27, many expressed concerns over slow progress on limiting emissions, including from the energy sector. As this year's Intergovernmental Panel on Climate Change report makes clear, the technology needed to support the transition exists today (Pathak et al., 2022). It has also fallen in price as it has advanced: for example, solar energy has gone from more than three times to less than two-thirds the price of coal in a decade (Pathak et al.). The financial case for renewables is now just as compelling as the environmental one and countries from Costa Rica to Scotland are increasingly turning to renewables. However, the key to unlocking the next level of progress lies elsewhere: human behaviour.

Understanding user behaviour and how people interact with energy systems can accelerate change – failing to do so could neutralise or even reverse hard-won technological advances. For example, in 2021 global energy consumption grew 4% as countries recovered from COVID-19 (International Energy Agency, 2021). In turn, many countries looked to non-renewable energy to plug short-term supply gaps, incurring long-term environmental costs. Consumer demand is the real driver of energy-related emissions – and therefore individual behaviour can drive change across the industry. In 2021 the Office for National Statistics found consumer expenditure was the largest contributor to UK emissions (27%; Office for National Statistics, 2022), while analysis shows at least 62% of emission reductions in the UK depend on behaviour change (Climate Change Committee, 2019). However, as our demand for energy falls, questions about whether it is produced from renewables can seem less important, potentially undermining the transition. So how can a focus on human behaviour support the shift?

1. Behaviour-centred design

The best design considers user behaviours and guards against human error – with green technology in particular, by reducing consumption. For example, we can anticipate the potential for using technology inefficiently – compromising its energy-saving potential – and mitigate accordingly. Using technology energy-efficiently should also 'feel better' than the alternative. By bringing together technical expertise and behavioural insights, we can motivate users to use less energy. The problem is complex: it is more energy efficient to only light occupied rooms, but people may feel safer in a brightly lit home. Good lighting design addresses both these competing concerns, such as using motion sensors which cut usage while maintaining or even increasing safety by warning of an intruder. Additionally, making more sustainable choices can be made easier using technology. Evidence suggests automation to control laptops, monitors, phones, and desk lights helped employees reduce energy use by up to 38% (Staddon et al., 2026), while occupancy-sensitive heating systems can cut consumption by up to a quarter (de Bakker et al., 2018).

2. Preventing rebound effects

It may seem counterintuitive but switching to renewables can *increase* energy consumption. Research shows that a 'rebound effect' appears when an increase in energy efficiency leads to less energy savings than expected, due to increased usage (Colmenares, Löschel & Madlener, 2020; Dutschke, Galvin & Brunzema, 2021; Lange et al., 2021). For example, after installing solar panels consumers may increase their energy usage because they feel they have license to because it is 'clean'; or after purchasing an electric vehicle, the consumer may take advantage of the lower running cost by driving more miles, more often.

The rebound effect has serious consequences: the cost of maintaining or replacing appliances, for example, and the more harmful possibility that this attitude spills over into activity which uses non-renewable energy, like driving a fossil fuel-powered vehicle. Long-term, this attitude may be passed down from parents to their children. This concept is known as 'moral licensing' and shows how we reward ourselves for our 'good' behaviour with moral credit we then use to pay for 'bad' elsewhere, relieving the discomfort of acting against our values or worldview (Burger, Schuler, & Eberling, 2022; Lasarov, Mai, & Hoffmann, 2022). As we saw in 2021, rising demand often precipitates a turn to non-renewable energy, incurring a carbon debt that future generations must repay. This can be tackled by focussing on a higher goal across all energy use – framed as 'cutting your energy footprint' and encouraging people to adopt a low-energy lifestyle.

3. Providing meaningful feedback

Behaviour change is a learning process: feedback reinforces new behaviour. We can influence this in a number of ways: for example, domestic smart meters which show households their exact energy consumption. However, feedback must be accessible and useful – after all, how many people can tell you what a kilowatt hour is? Research shows that seeing consumption in financial terms is more powerful in reducing household energy use – by around 10% (Darby, 2006). The social dimension is also key: allowing users to compare their consumption to their neighbours and combining smart meters with goal-setting – a powerful driver of behaviour change – could cut consumption even more. Studies in the workplace have shown highlighting and rewarding energy saving behaviour can cut consumption by up to 12% (Staddon et al., 2016).

4. Understanding how behaviour scales

For the first time in human history, we can see how individual behaviour scaled across the lifetimes of billions of people has global consequences. Sustained, widespread energy efficiency happens through the accumulation of the many choices people make every day. Leaving a light on in an unoccupied room appears trivial – but how many unoccupied rooms are needlessly lit right now? Scale across a year and it is clear how seemingly small changes accumulate to create demand which remains largely satisfied through fossil fuels. Some sports teams are switching off stadium floodlights straight after a match – this might not scratch the surface yet, but what if entire leagues held fixtures an hour earlier to reduce energy consumption throughout the year? We don't perceive this intuitively: one way to help is to give consumers feedback. Several studies have shown that instant feedback via digital systems like online dashboards is effective at encouraging people to use less energy (Darby, 2016; Fischer, 2008; Yun et al., 2013). This is because this can help people evaluate their behaviour and encourage them to adapt. Government can also encourage collective change in the long-term (such as maintaining short-term reductions in consumption due to higher energy bills).

Conclusion

Centuries of human development ran on what we now know was highly-polluting energy – so shifting away from fossil fuels is a fundamental transformation. Energy consumers (or, to put it another way, every single one of us) have a key role to play in changing how we live our daily lives. Developing new technologies is critical in tackling the climate crisis but we must deploy it effectively. Human behaviour sits at the crossroads between people and technology: systems must be designed to consider the complex, changing relationship between the two. Without considering how technology helps consumers reduce their carbon footprint, we risk investing in solutions which don't deliver the change we need. The scale of the problem, the urgent need to address it and the fundamental nature of energy means neglecting this will have dire consequences – and, therefore, even more radical changes in our lives. Reducing the energy consumption we control can drive this all-important transition – otherwise our efforts to tackle climate change could be in vain.

References

Burger, A. M., Schuler, J., & Eberling E. (2022). Guilty pleasures: Moral licensing in climate-related behavior. *Global Environmental Change, 72*, 102415. https://doi.org/10.1016/j.gloenvcha.2021.102415

Climate Change Committee (2019). *Net Zero: The UK's contribution to stopping global warming*. Climate Change Committee, United Kingdom. Retrieved from https://www.theccc.org.uk/publication/net-zero-the-uks-contribution-to-stopping-global-warming/

Colmenares, G., Löschel, A., & Madlener, R. (2020). The rebound effect representation in climate and energy models. *Environmental Research Letters, 15*, 123010. https://doi.org/10.1088/1748-9326/abc214

Darby, S. (2006). *The effectiveness of feedback on energy consumption: A review for DEFRA of the literature on metering, billing and direct displays.* Environmental Change Institute, Oxford University.

de Bakker, C., Aarts, M.P., Kort, H., & Rosemann, A. (2018). The feasibility of highly granular lighting control in open-plan offices: Exploring the comfort and energy saving potential. *Building Environment*, *142*, 427–438.

Dutschke, E., Galvin, R., & Brunzema, I. (2021). Rebound and spillovers: Prosumers in transition. *Frontiers in Psychology, 12,* 636109. https://doi.org/10.3389/fpsyg.2021.636109

Fischer, C. (2008). Feedback on household electricity consumption. *Energy Efficiency, 1*, 79–104.

International Energy Agency (2021). *Global energy review 2021: Assessing the effects of economic recoveries on global energy demand and CO2 emissions in 2021.* International Energy Agency, Paris: OECD.

Lasarov, W., Mai, R., & Hoffmann, S. (2021). The backfire effect of sustainable social cues: New evidence on social moral licensing. *Ecological Economics, 195*, 107376. https://doi.org/10.1016/j.ecolecon.2022.107376

Lange, S., Kern, F., Peuckert, J., & Santarius, T. (2021). The Jevons paradox unravelled: A multi-level typology of rebound effects and mechanisms. *Energy Research and Social Science, 74*, 101982. https://doi.org/10.1016/j.erss.2021.101982

Office for National Statistics (2022) *Greenhouse gas emissions, UK: Provisional estimates 2021*. Retrieved from https://www.ons.gov.uk/economy/environmentalaccounts/bulletins/greenhousegasintensityprovisionalestimatesuk/2021

Pathak, M., Slade, R., Shukla, P. R., Skea, J., Pichs-Madruga, R., Ürge-Vorsatz, D. (2022). Technical Summary, in: Climate Change 2022: Mitigation of Climate Change. *Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (eds Shukla, PR et al.)* Cambridge University Press, Cambridge, UK.

Staddon, S., Cycil, C., Goulden, M., Leygue, C., & Spence, A. (2016). Intervening to change behaviour and save energy in the workplace: A systematic review of available evidence. *Energy Research and Social Science*. *17*, 30-51.

Yun, R., Lasternas, B., Aziz, A., Loftness, V., Scupelli, P., Rowe, A., Kothari, R., Marion, F., & Zhao, J. (2013). Toward the design of a dashboard to promote environmentally sustainable behavior among office workers. In *Persuasive Technology: 8th International Conference, PERSUASIVE 2013, Sydney, NSW, Australia, April 3-5, 2013. Proceedings 8* (pp. 246-252). Springer Berlin Heidelberg.

Lesson Learned: the similarities and differences of human factors in Aircraft Maintenance between JL123 and CI611

Punthit Kulsomboon[1], Edem Yao Tsei[1], Gayatri Rebbapragada[1] & Wen-Chin Li[1]

[1]Safety and Human Factors in Aviation MSc, SATM Cranfield University, United Kingdom

SUMMARY

This paper makes use of the human factors analysis and classification system (HFACS) to analyse maintenance-related causal factors of two accidents - Japan Air Lines (JAL) flight 123 (JL123) and China Airlines (CAL) flight 611 (CI611). Furthermore, the pathways that could have resulted in the two accidents were identified by applying the HFACS framework. The study also compares the similarities and differences between these accidents. The findings of this paper lend support to past research on HFACS where higher levels at an organisation have been shown to have directly affected the lower levels. Lessons from these accidents have also been identified in order to prevent recurrences.

KEYWORDS

Active failure, Human Factors Analysis and Classification System, Latent Failure

Introduction

Over the years, there has been a shift in the human factors focus on determining the causes of aviation accidents during investigations. The current focus includes decision-making, supervisory factors and organisational culture among others as compared to the earlier focus on skill deficiencies. The human factors analysis and classification system (HFACS) framework by Shappell and Wiegmann (2001) allows investigators to focus on these factors and it is based on Reason's organisationally based model of human error (Reason, 1990). This framework has four levels which are level 1 (unsafe acts of operators – active failures), level 2 (preconditions for unsafe acts – latent and active failures), level 3 (unsafe supervision – latent failures) and level 4 (organisational influences- latent failures). The relevance of the HFACS framework in human factors can be seen in its modern applications in human reliability assessment for complex space operations (Alexander, 2019). The framework was also recently applied as a proactive prevention in public health during COVID-19 (Bickley & Torgler, 2021) and most relevant to this paper is the adaptation of the framework to aircraft maintenance deviations (Illankoon et al., 2019). The HFACS framework is depicted in Figure 1.

Past research has concluded that there is a relationship between the errors that occur at lower levels and inadequacies at higher levels in an organisation (Li & Harris, 2013). The accident of JL123 saw the rupture of the pressure bulkhead that led to a loss of flight control that subsequently led to the crash of the aircraft (Aircraft Accident Investigation Commission, 1987). The accident of CI611, on the other hand, involved the in-flight break-up of the aircraft as it approached its cruising altitude (Aviation Safety Council, 2002). Problems with maintenance have been considered as having had the most significant impact which allowed these two accidents to occur (Jiang, 2020). Thus, the aim of this paper is to analyse and compare the maintenance-related causal factors of JL123 and CI611 accidents using the HFACS framework in order to learn lessons and prevent recurrences.

Methodology

This is a qualitative study of applying the HFACS framework which consists of 18 categories to analyse the JL123 and CI611 accident reports to determine causal factors related to maintenance. Four aviation human factors researchers formed a subject matter expert focus group and conducted a content analysis based on the accident reports. The analysts had received detailed training on the HFACS framework. The presence (coded 1) or absence (coded 0) of each HFACS category was evaluated from the narrative of each accident report. Where there were discrepancies in the categorisation of an accident, the researchers convened and resolved differences in observations.

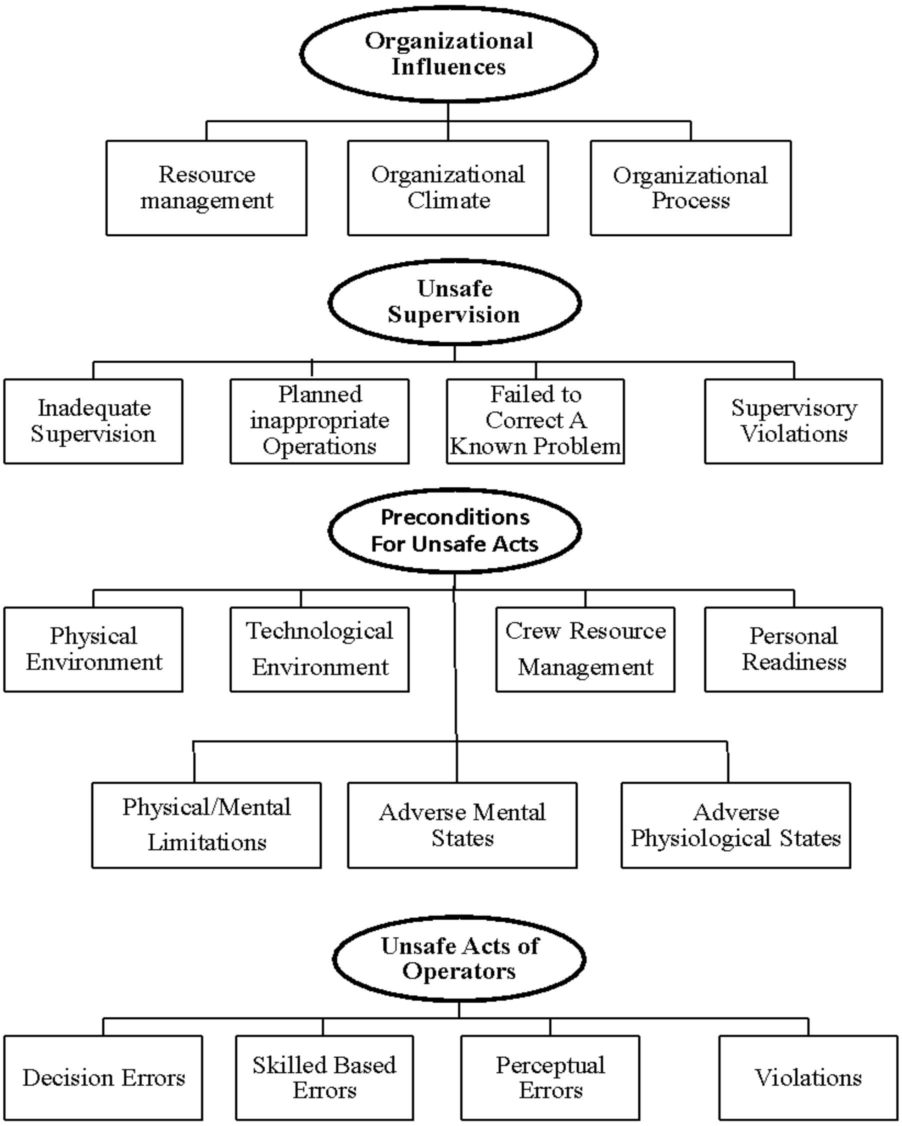


Figure 1: The HFACS framework- Shappell and Wiegmann (2001)

Results and Discussion

The two accidents were analysed and discussed in the focus group and the similarities and differences based on the executive summaries of official accident reports for both aircraft were identified.

Similarities and differences between CI611 and JL123 Accidents

The similarities between the two accidents are straightforward. In both instances, the aircraft involved were Boeing 747 models and the aircraft had suffered a tail strike prior to the actual accidents. Moreover, the maintenance work done after the tail strike was not according to the Boeing Structural Repair Manual (SRM). The lack of quality follow-up-maintenance inspections, post-repairs, was also observed in both instances. This may have resulted in maintenance personnel in both cases not detecting cracks from metal fatigue following the repair works. Furthermore, both airlines during the planning phase had aimed at carrying out proper repair works, however, the actual repair works done were different from the intended corrective measures. Lastly, maintenance records were found to be incomplete in both cases.

The series of aircraft models involved in both accidents were different. In the case of CI611, repair works on the bulkhead after a tail strike incident was conducted internally by CAL Engineering Maintenance Division (EMD). On the other hand, JL123 repair works on the tail were conducted by a Boeing AOG (Aircraft on Ground) repair team dispatched by Boeing and contracted by JAL. The accident of CI611 resulted in the disintegration of the body of the aircraft into different parts, whereas flight JL123 had only lost its tail prior to crashing. Flight radar data indicated that the JL123 bulkhead raptured at about twenty-four thousand feet whilst the complete disintegration of CI611 occurred at a higher altitude of about thirty-four thousand nine hundred feet above mean sea level. Additional information on the comparison can be found in Table 1.

Table 1: Comparison between JL123 and CI611

Flight	JL123	CI611
Type of Aircraft	Boeing 747 SR-100	Boeing 747-200
Manufactured Date	30 January 1974	15 July 1979
Total Flight Hours	25,030	64,810
Prior Tail Strike Incident Date	2 June 1978	7 February 1980
Accident Date	12 August 1985	25 May 2002
Year interval between Tail Strike and Crash	7 years	22 years
Number of Casualties	520	225
Total Number of landings	18,835	21,180

The rupture of the tail bulkhead of JL123 destroyed the hydraulic lines which controlled the pitching and yawing movements of the aircraft; this in turn led to the loss of flight controls. Flight characteristics of JL123 made it difficult for pilots to control. The captain and first officer (FO), however, put in significant efforts by combining available crew resources (use of power levers and aileron controls) to control the aircraft. In spite of all efforts of the flight crew, JL123 eventually crashed (Aircraft Accident Investigation Commission, 1987). Such a demonstration of Crew Resource Management (CRM) needs to be commended given the fact that more than half of events involving loss of flight control result in an accident and half of these accidents result in a catastrophe (Jacobson, 2010). The timely and effective CRM applied moments before the crash of JL123 may have allowed those four individuals to survive such a tragedy.

Analysis using HFACS framework

The focus group also involved categorising each of the causal factors from JL123 (three points) and CI611 (six points) as listed in the executive summary using the HFACS framework. The causal factors were discussed to determine the pathway between the HFACS categories that might have been followed and resulted in the accidents. All pathways for JL123 (in blue) and CI611 (in red) identified in the focus group are in Figure 2. The solid lines represent the direct relation between a high level and their immediate lower level on the HFACS framework. Furthermore, the dotted lines represent a direct relationship between higher levels of HFACS (level 3 and level 4) and lower levels (level 1 and level 2). The dotted lines thus denote that actions at the highest level can directly impact the lowest levels without necessarily interacting with the levels in between. Solid black boxes bring focus to the categories of the HFACS that were found to be relevant for the two accidents. On the other hand, the grey dotted boxes represent categories found to be unrelated to the two accidents.

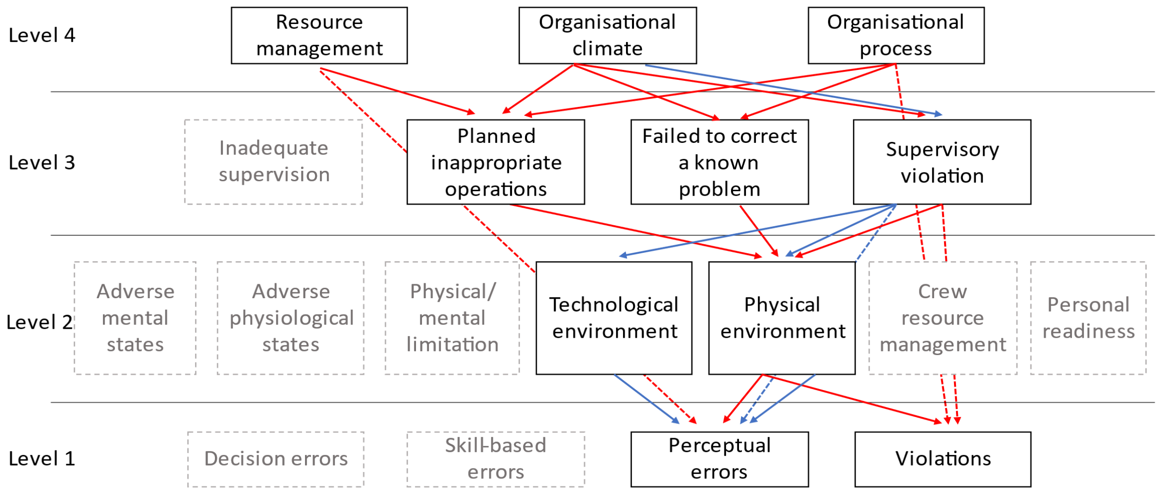


Figure 2: Pathway between the HFACS categories for JL123 (blue) and CI611 (red)

Impact pathway for JL123 accident

The impact pathway for JL123, as determined by the participants, is depicted using the colour blue in figure 2. For JL123, the organisational climate (level 4), such as poor safety culture, could have led to supervisory violations (level 3) during the repair works. Such violations could have resulted in improper repairs being carried out after the tailstrike incident. Additionally, supervisory violations may have also contributed to the lack of proper checks during subsequent inspections in the years following the tailstrike incident. Such supervisory violations during the repair might have led to a deterioration of the technological environment (level 2) during the flight in the form of the loss of primary flight controls.

Moreover, it could have been that failures at the higher levels could have had a direct impact on the failures at the lower levels for the JL123 accident. These pathways are represented by dotted blue lines in figure 2. In the case of JL123, it could have been that supervisory violations (level 3) in the form of improper repair work carried out under the direction of the supervisors might have directly led to perceptual errors (level 1) in the form of the maintenance staff's inability to detect cracks during the inspection.

Impact pathway for CI611 accident

The impact pathway for CI611, as determined by the participants, is depicted using the colour red in figure 2. For CI611, the possible poor organisational climate (level 4) at CAL and its poor resource management (level 4), such as not providing magnifying glasses and lighting during the inspection, could have led to supervisory violation (level 3) in the form of improper maintenance operations which in turn could have impacted the physical environment (level 2) of the maintenance team, for example, poor lighting condition. These findings, therefore, are similar to the findings of Li & Harris (2013) given that in both accidents, factors at lower levels like perceptual errors and physical environment are impacted by higher levels like supervisory violation and poor organisational climate.

Just like for JL123, in the case of CI611, there were also possible situations where failures at higher levels could have directly impacted the lower levels. These pathways are represented by dotted red lines in figure 2. In the case of CI611, it was suspected that violations and perceptual errors (level 1) made by maintenance staff like not carrying out the right repair following the tail strike incident and failing to observe the impact of the wrong repairs during subsequent inspections could have been direct consequences of organisational process and resource management (level 4) respectively. Poor resource management (level 4) could be depicted, for example, by the inability of CAL to provide eddy-current detection tools to conduct a non-destructive test during subsequent inspections.

Participants also deemed that the catastrophic outcome of the CI611 accident could have been avoided had the safety culture at CAL been more positive. Another possible explanation of the impact pathway of the accident can be explained with regards to the safety culture. According to Li and Harris (2006), national culture affects the safety culture of an organisation. The latter, in turn, is heavily dependent on the actions of senior management. The lack of suitable actions from the senior management with regards to safety at CAL has already been previously identified. It is worth noting that the aircraft involved in the accident was allowed to remain in operation even with evident safety issues (following the tailstrike incident) since 1997. This means the organisation had at least five years to make corrections to the improper repairs had inspections been carried out properly. This brings into focus the attitudes of the senior management which could have negatively impacted the safety culture at CAL. Such a safety culture could have impacted resource management, organisational climate, and organisational process (level 4). This, in turn, could have led to failures at level 3 like planned inappropriate operations, failed to correct a known problem, and supervisory violation. All of which could have been a result of the poor or lack of quality supervision from middle-level management. Failures at level 3 could be the failure to carry out the appropriate repairs and delays in carrying out the said repairs. The lack of supervision (another failure at level 3) might have also impacted the physical environment (level 2) for the maintenance team carrying out the supervision in the form of poor lighting and lack of magnifying glasses.

Similarities between the impact pathways of JL123 and CI611 accidents

The maintenance supervisors are responsible to provide their personnel with resources, facilities and a working environment to succeed and ensure repairs are done safely and efficiently (Harris & Li, 2011). It was deemed that for the maintenance supervisors at JAL (Boeing maintenance company) and those at CAL (CAL EMD), these qualities were absent and this was a possible supervisory violation. Such a supervisory violation (level 3) may have affected the performance of the maintenance engineers carrying out the repairs and the maintenance environment generally that could have allowed perceptual errors (level 1) during repair works in the case of JL123 (represented by the blue dotted line) and violations of Boeing SRM in the case of CI611 (represented by red dotted lines).

Interestingly, for both JL123 and CI611 accidents, no pathways connecting adverse mental states, or physical/mental limitations (level 2) were found. This means that the participants did not classify these categories as having contributed to the occurrence of the two accidents. The findings of this analysis are dependent upon the causal factors identified in the official reports. It may have been that the accident reports did not find any such psychological precursors at the management and supervisory level of maintenance as contributing factors to the accidents.

Differences between the impact pathways of JL123 and CI611 accidents

In the case of JL123 only one category, that is, organisational climate from level 4 seems to have played a role in the occurrence of this accident. It could be that the organisational climate at JAL disregarded or did not prioritise safety. This could have been the precursor for supervisory violation (level 3) to take place that allowed improper repairs to be made. Supervisory violation was the only category from level 3 that was identified as part of the HFACS analysis. Subsequent pathways that may have allowed the accident to occur have been described above. On the other hand, in the case of CI611, all of the categories from level 4, which are poor resource management, poor organisational climate, and poor organisational process, seemed to have impacted three of the four categories at level 3, which included supervisory violation, planned inappropriate operations and failure to correct a known problem. It could be that the poor organisational culture may have affected numerous levels within CAL. A lack of appropriate response after the tailstrike from the top management at CAL (failure to correct a known problem) might have resulted in delayed repair works and the planning of inappropriate maintenance works which did not follow the procedures as recommended by Boeing (planned inappropriate operations) both of which could have been a result of poor supervision. Subsequent pathways that may have allowed the accident to occur have been described above.

Additionally, in the case of the CI611 accident, poor resource management (level 4) like not providing the maintenance teams with the appropriate resources (like magnifying glasses and proper lighting) might have directly led to perceptual errors (level 1) by the maintenance teams. Due to the unavailability of proper resources, it could be that the maintenance teams were unable to properly carry out inspections following the tailstrike incident. Furthermore, the organisation process (level 4) perhaps not taking immediate action following the tailstrike incident and delaying repair work could have resulted in violations (level 1) in the form of the incorrect repair being done after the tailstrike incident. This pathway is represented in red dotted lines in figure 2. No impact of level 4 directly on level 1 was observed in the case of the JL123 accident.

Conclusion

The aviation industry can learn lessons from the accidents of JL123 and CI611. Based on the findings of this study, for there to be a significant improvement in the overall safety of aviation maintenance, interventions must primarily focus on level 3 and level 4. This study identified “organisational climate”', “resource management” (level 4) and “supervisory violation” (level 3) that impacted lower levels in both cases. Additionally, the study also highlighted the importance of national culture and its impact on CAL’s organisational safety culture which significantly contributed to the crash of CI611. It is thus possible to conclude that to avoid recurrences, the culture of the country and the airline will need to be one that prioritises safety and takes timely actions. The development of a good safety culture needs to be a top down process. For the safety culture to succeed, it is imperative that all shareholders like the senior management, the middle management, and the lower-level operators within an airline are involved. Continuous monitoring in the form of regular safety inspections and audits must take place. Proper records of these inspections must be maintained for future reference. Findings of these inspections and audits must be used to continuously improve the safety standards within an airline. Lastly, the regulators must work towards improving the safety standards by providing airlines with proper inspection training and by keeping a watchful eye on the operations carried out by the airlines, especially those that raise safety-related concerns.

This study lends support to the theoretical basis of the HFACS framework and past research. Future accidents can be avoided by targeting changes in the areas identified above. Effective human factor interventions need to be introduced to prevent the recurrence of such accidents.

References

Aircraft Accident Investigation Commission. (1987). Aircraft Accident Investigation Report Japan Air Lines Co., Ltd. Aircraft Accident Investigation Commission Ministry of Transport, Tokyo, Japan

Alexander, T. M. (2019). A case based human reliability assessment using HFACS for complex space operations. In Journal of Space Safety Engineering (Vol. 6, Issue 1, pp. 53–59). Elsevier Ltd. https://doi.org/10.1016/j.jsse.2019.01.001

Aviation Safety Council. (2002). In-flight breakup over Taiwan Strait Northeast of Makung, Penghu Island China Airlines Flight CI-611. Report no. ASCAOR-05-02-001. Aviation Safety Council, Taipei, ROC.

Bickley, S. J., & Torgler, B. (2021). A systematic approach to public health – Novel application of the human factors analysis and classification system to public health and COVID-19. Safety Science, 140. https://doi.org/10.1016/j.ssci.2021.105312

Harris, D., & Li, W.-C. (2011). An extension of the Human Factors Analysis and Classification System for use in open systems. Theoretical Issues in Ergonomics Science, 12(2), 108–128. https://doi.org/10.1080/14639220903536559

Illankoon, P., Tretten, P., & Kumar, U. (2019). A prospective study of maintenance deviations using HFACS-ME. International Journal of Industrial Ergonomics, 74. https://doi.org/10.1016/j.ergon.2019.102852

Jacobson, S. (2010). Aircraft Loss of Control Causal Factors and Technical Challenges, American Institute of Aeronautics and Astronautics Conference - Final.doc. Ntrs. https://ntrs.nasa.gov/api/citations/20100039467/downloads/20100039467.pdf

Jiang, T. W. (2020). An Inductive Study of Aviation Maintenance Human Errors and Risk Controls. https://doi.org/10.13140/RG.2.2.19884.28808

Li, W.-C. & Harris, D., (2006). Where safety culture meets national culture: the how and why of the China Airlines CI-611 accident. Human Factors and Aerospace Safety, 5(4), 345–353.

Li, W.-C., & Harris, D. (2013). The identification of training deficiencies in pilots by applying the human factors analysis and classification system. International Journal of Occupational Safety and Ergonomics, 19(1), 1-16. doi:10.1080/10803548.2013.11076962.

Reason, J.T., (1990). Human Error. Cambridge University Press, Cambridge.

Shappell, S.A., & Wiegmann, D.A. (2001). Applying reason: the human factors analysis and classification system (HFACS). Hum. Factors Aerospace Saf. 1, 59–86

Shappel, S.A., & Wiegmann, D.A. (2003), A Human Error Analysis of General Aviation Controlled Flight Into Terrain Accidents Occurring Between 1990-1998, Report no. DOT/FAA/AM-03/4 (Washington, DC: Federal Aviation Administration).

Ghost Busting: A Novel On-Road Exploration of External HMIs for Autonomous Vehicles

David R. Large, Madeline Hallewell, Xuekun Li, Catherine Harvey, Gary Burnett

Human Factors Research Group, University of Nottingham

SUMMARY

The absence of a human driver in future autonomous vehicles means that explicit pedestrian-driver communication is not possible. Building on the novel 'Ghost Driver' methodology to emulate an autonomous vehicle, we developed prototype external human-machine interfaces to replace existing cues, and report preliminary, qualitative findings captured from a sample of pedestrians (n=64) who encountered the vehicle when crossing the road, as well as reflecting on the method.

KEYWORDS

Autonomous vehicles, external HMI, Ghost Driver

Introduction

There has been considerable interest amongst behavioural scientists in the potential impact of highly and fully autonomous 'self-driving' vehicles (AVs) on the behaviour of pedestrians. These vehicles, operating at SAE level 4 or 5, are unlikely to have a human driver present, and as such, explicit visual cues (head, eye, hand/arm gestures etc.) that are traditionally exchanged between a driver and a pedestrian, will be absent. Typically, these aim to establish a mutual understanding of perception (*Have you seen me?*) and intent (*Will you give way?*) (Merat et al., 2018), and are important to overall traffic safety especially in low-speed crossing scenarios in complex urban settings (Lee et al., 2020). However, studying genuine, naturalistic behaviours of people responding to AVs presents a number of challenges (limited public trials, requirement to have a 'safety driver' present etc.). A novel solution is to use a Wizard-of-Oz (WoZ) approach to give the appearance that the car is driving on its own, even when it is not. This can be achieved by hiding the driver using a bespoke seat cover (aka 'Ghost Driver' method) (Rothenbücher et al., 2016). To date, no such studies have been reported in the UK. In addition, the Ghost Driver method has not been employed specifically to evaluate external human-machine interfaces (eHMIs).

Method

A 'Ghost Driver' WoZ study was devised in which the driver was hidden in a bespoke seat-suit, thereby giving the appearance that the vehicle (Nissan Leaf) was driving by itself (Figure 1). The seat-suit was designed and fabricated to enable the driver to maintain safe control of the vehicle, whilst ensuring that they could not be seen by a passing pedestrian glancing into the vehicle. Three eHMIs were created. These were informed by the literature and prototyped using an individually addressable RGB-LED matrix and strip attached to the outside of the vehicle (front of bonnet and at top of windscreen, respectively). The eHMIs were programmed using an Arduino Mega board and manipulated with push-button controls from inside the vehicle. The eHMI designs employed varying degrees of anthropomorphism (implicit, explicit, low) to aid interpretation and build trust. The first (implicit) utilised the LED strip only and mimicked the pupillary response of an eye: lateral movement demonstrated scanning/awareness, and blinking provided an implicit cue of the

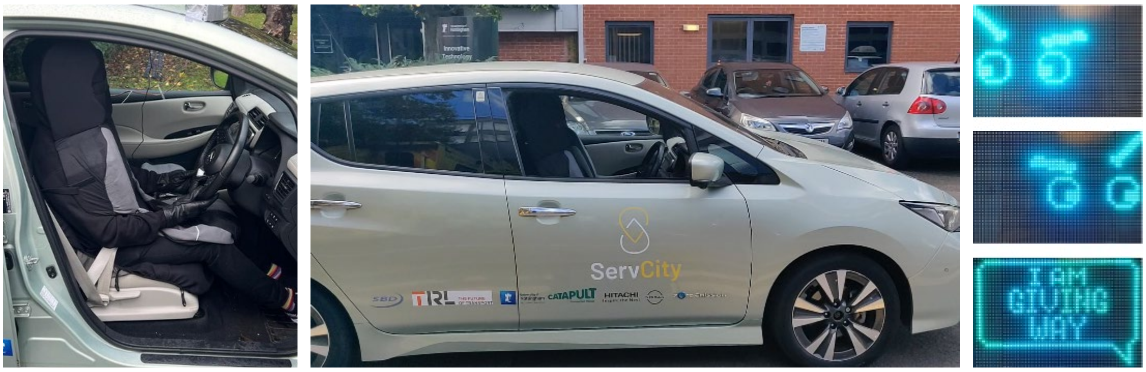


Figure 1: Driver in seat-suit (left); hidden driver operating car (centre); example eHMIs (right)

vehicle's intention to give way. The second (explicit) presented a face and eyes on the LED matrix to scan the road and used humanlike language to 'talk' to the pedestrians (Figure 1). The third (low) used a vehicle icon and vehicle-centric language on the LED matrix. For each eHMI design, four modes were created: scanning, giving way (pedestrian on right), giving way (pedestrian on left) and giving way (pedestrians on both sides of road). A second researcher, seated in the back seat of the ego-vehicle, controlled the current state of the eHMI in response to the observable pedestrians in the vicinity of the vehicle. The study took place on the extensive University of Nottingham campus and a circuitous route was selected that included several marked and unmarked crossings. Over 10 hours of video data were captured using a dashcam and GoPro recorder to document pedestrians' responses to the 'driverless' vehicle and eHMIs. In addition, researchers were located at specific crossing points, and invited pedestrians who encountered the vehicle to complete a survey.

Results and Discussion

Video analysis is ongoing. Here, we report qualitative findings, including illustrative comments and responses related to the vehicle and eHMI concepts. Results show that over eighty percent of respondents believed that the car was driving on its own ("*There was no driver, just a passenger in the back passenger seat*"), and this surprised many people ("*I was mostly just shocked, so I stopped and observed*"). Nevertheless, many people still appeared to interact with the vehicle as if a driver were present (e.g. waving to thank the vehicle for stopping), highlighting the value of an eHMI to replace interactions with a driver, and supporting the inclusion of 'human' elements. Comments suggest the eHMIs impacted the trust relationship ("*I was a bit curious about why the car stopped...when I saw the screen that explained a lot*"), with most comments suggesting support for the concepts ("*I quickly became aware that it was helping me to cross*", "*[the eHMI] matched observed behaviour of vehicle*", "*I understood that the eyes were looking out for people*"), whereas others were more cautious towards the technology ("*Would need to encounter it more before I fully trusted it*"), and a few respondents admitted being confused by the messages ("*I wasn't entirely sure what the message was conveying*"). This did not necessarily change pedestrians' crossing behaviour, with most respondents still stating that they crossed in front of the vehicle as they normally would. It did, however, inspire some additional curiosity: "*Had seen it…earlier and was curious to see if it would stop or not.*" The different eHMIs appeared to inspire different emotional responses. For example, the explicit anthropomorphism encouraged positivity – smiling, laughing etc., whereas responses to the low anthropomorphism were more perfunctory; survey ratings indicated that the latter provided the highest clarity in conveying its intended messages. Overall, initial findings support the use of a hidden 'ghost driver' to explore pedestrians' interactions with an AV, with observed behaviour suggesting high ecological validity. In addition, explicit communication using eHMIs (employing elements of anthropomorphism) appears to encourage safe crossing behaviours, help pedestrians interpret vehicle behaviour and intent, and increase their confidence and build appropriate trust when interacting with a driverless vehicle.

Acknowledgements

The study was undertaken as part of the ServCity project (https://www.servcity.co.uk/), which is funded by the UK Innovation Agency (Innovate UK) and Centre for Connected & Autonomous Vehicles (CCAV) (Grant number: 105091). We would like to thank Nissan Technical Centre Europe (NTCE) for use of the Nissan Leaf vehicle and their support and guidance throughout.

References

Lee, Y. M., Madigan, R., Giles, O., Garach-Morcillo, L., Markkula, G., Fox, C., . . . Merat, N. (2020). Road users rarely use explicit communication when interacting in today's traffic: implications for automated vehicles. Cognition, Technology and Work.

Merat, N., Lee, Y. M., Markkula, G., Uttley, J., Camara, F., Fox, C., . . . Schieben, A. (2018). How do we study pedestrian interaction with automated vehicles? Preliminary findings from the European interACT project. In G. Meyer & S. Beiker (Eds.), Lecture Notes in Mobility (pp. 21-33). Cham: Springer.

Rothenbücher, D., Li, J., Sirkin, D., Mok, B., & Ju, W. (2016). Ghost driver: A field study investigating the interaction between pedestrians and driverless vehicles. Paper presented at the 2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN).

Assessing pilots' situation awareness using touchscreen inceptor

Kyle Hu[1], Wojciech Tomasz Korek[1&2] & Wen-Chin Li[1]

[1] Safety and Human Factors in Aviation MSc, SATM, Cranfield University, United Kingdom
[2] Faculty of Automatic Control, Electronics & Computer Science, Silesian University of Technology, Poland

SUMMARY

This research aims to investigate the innovative technology of touchscreen inceptor impact on pilots' situation awareness compared to a traditional sidestick. The Pupil Lab eye tracker collected pilots' fixation counts in the Future Systems Simulator (FSS), and subjective measure was the situation awareness rating technique (SART). A significant difference was spotted in the attention demand, understanding, and total score from SART analysis. Furthermore, the visual parameter of fixation counts indicated that pilots spent less time on OTW and more time on PFD when interacting with the touchscreen inceptor compared to the sidestick. The findings show the potential to implement a touchscreen in future flight deck designs.

KEYWORDS

Flight Deck Design, Human-Computer Interaction, Situation Awareness, Touchscreen Inceptor

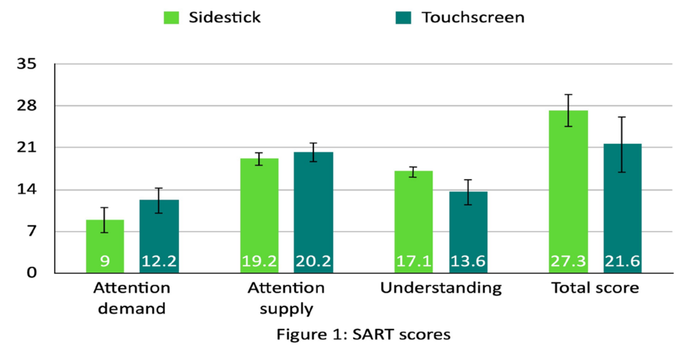
Introduction

Effective hand-eye coordination and no extra space required features made touchscreens one of the most welcomed technologies in daily use. Technology can mimic the mechanical components via the display and present the potential for reducing pilots' task loads by interacting with practical human-computer interface design (Korek et al., 2022). While interacting with a touchscreen, human operators were not required to fixate on other displays to search for the consequences of their inputs compared with traditional mechanical knobs/levers on flight decks. The replacement of touchscreens has the potential to reduce training time and cost; however, its impact on aircrew's performance must be thoroughly examined before introduction (van Zon et al., 2020). Therefore, this study aims to extend the potential of using the touchscreen as an inceptor to explore the potential issues of human-computer interaction in the future flight deck.

Methods

Ten participants aged 22 to 46 (M = 29.6, SD = 7.8) with flying experienced (M = 695.7, SD = 1001.7) were involved in the experiment. The future systems simulator (FSS) was utilized, as it allowed the use of both touchscreen and sidestick flight inceptors on instrument landing scenarios. Participants were asked to conduct an instrument landing using a sidestick and touchscreen inceptor wearing an eye-tracking device. Automation systems controlled the speed and rudder to simulate the highly automated flight deck and mitigated deviation. There are three areas of interest in the FSS, including "out of the window" (OTW) view, navigation display (ND), and the primary flight display (PFD). Pilots' situation awareness was measured with SART, which consists of ten questions in three dimensions (supply, demand, and understanding) from zero (low) to seven (high) (Taylor, 2017).

Result



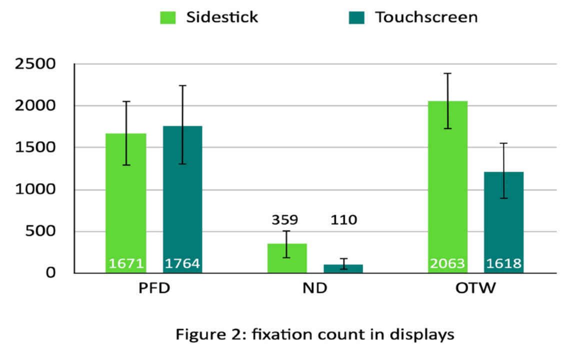


Figure 1: SART results

The result of SART demonstrated that attention demand is significant, t(10) = 3.28, p = 0.04, d = 1.04, and significance on understanding, t(10) = 3.41, p = 0.003, d = 1.08. There is no significance on attention supply, t(10) = 0.73, p = 0.24. However, there is a significance in SART total score, t(10) = 2.37, p = 0.02, d = 0.75 (figure 1). The visual parameters demonstrated no significance on fixation counts on PFD, t(10) = 0.47, p = 0.32. However, pilots' fixation counts were significantly less on both ND, t(10) = 2.04, p= 0.04, d = 0.65, and OTW, t(10) = 4.47, p = 0.0008, d = 1.41 while interacting with a touchscreen compared with sidestick (figure 2).

Discussion & Conclusion

When a touchscreen is used as an inceptor to manipulate the aircraft landing, the system demand on the attentional resources from pilots and supply of attentional resources to pilots is higher than sidestick. However, pilots' understanding of the situation could have been higher, possibly due to the innovative functions of touch-control overlapping with PFD with critical information, such as airspeed, altitude, heading and glideslope. Therefore, pilots' total SART scores were significantly lower on touchscreen interaction than on sidestick (figure 1). Furthermore, pilots' visual attention on OTW and ND showed significantly fewer fixation counts on the touchscreen. The significantly reduced fixation counts on OTW indicated that pilots head-down time significantly increased while interacting with the touchscreen. This may negatively impact pilots' SA performance, as pilots could not perceive the dynamic changes of the surrounding environment. Although the fixation counts on the PFD did not show a significant difference (figure 2), their fingers blocked the critical information related to airspeed and altitude while interacting with the touchscreen. Changing the touchscreen layout may be a solution, for example, by installing the touchscreen inceptor in a suitable place, which can eliminate the operational risk of obstructing critical information.

Touchscreen inceptors may provide potential benefits for further development, though their application is still in the infancy stage and is not yet ready for implementation in the flight deck. This study suggests that introducing a touchscreen as an inceptor in a flight deck needed further consideration of potential human factors issues. Human-computer interactions and user interface design can help the technology be more suitable for human-centred design and enhance the chance of successfully integrating touchscreen inceptors into future flight decks.

Acknowledgments

This research is funded by the European Union through the European Social Fund (POWR.03.02.00-00-I029). The authors would like to thank the FSS Team in Cranfield, especially Mudassir Lone, for his generous support during the project’s development, and Rolls-Royce, particularly Peter Beecroft, for approving the research to be carried out in the Future Systems Simulator.

References

Korek, W. T., Li, W. C., Lu, L., & Lone, M. (2022). Investigating Pilots' Operational Behaviors While Interacting with Different Types of Inceptors. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *13307 LNAI*, 314–325.

Taylor, R. M. (2017). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. *Situational Awareness*, 111–127.

van Zon, N. C. M., Borst, C., Pool, D. M., & van Paassen, M. M. (2020). Touchscreens for Aircraft Navigation Tasks: Comparing Accuracy and Throughput of Three Flight Deck Interfaces Using Fitts' Law. *Human factors*, *62*(6), 897–908. https://doi.org/10.1177/0018720819862146

Exploratory study of virtual reality flight training device for upset prevention and recovery training

Filip Florek

Cranfield University, UK

SUMMARY

This exploratory study aimed to establish whether a virtual reality (VR) enabled flight training device will provide effective Upset Prevention and Recovery Training (UPRT) to ensure equivalent safety with the Flight Simulation Training Device (FSTD) considering pilot situational awareness competency. It was achieved by determining the effects of the pilot's presence, task-related stress and cybersickness on situational awareness during upset prevention, and, if necessary, recovery and by assessing pilot acceptance of VR-enabled flight training device in UPRT. No evidence has been found that situational awareness was negatively affected by exposure to VR, with certain reactions to stimuli degradation, i.e., flight upset in Instrument Meteorological Conditions (IMC) resulting in a predictable outcome of increased attentional demand.

KEYWORDS

Serious Games for Training, Simulation, Situation awareness, Virtual Reality, Training.

Introduction

Many sources (Airbus, 2022; Boeing, 2021) cite Loss of Control In-flight (LOC-I) as a leading causal factor in fatal accidents in civil aviation over the last 30 years. In recent decades technological advancements like flight envelope protection and advanced upset warning systems allowed to reduce LOC-I fatal accident rates by nearly 90% (Airbus, 2022), but still high-profile accidents like Air France (Flight 447) and Colgan Air (Flight 3407) undermined the mitigating factors and proved that that flight automation alone will not resolve the issue (Richards et al., 2012). In both cases, the accident report articulates the complexity of contributing factors leading to airplane aerodynamic stall and pilot's cognitive performance degradation, especially in situational awareness and as result inadequate response by pilots (BEA, 2012; NTSB, 2009).

Aircraft upset describes an in-flight state in which an aircraft exceeds structural parameters of the airframe (ICAO, 2014b). These disturbances may result in a stall, spin, or over-limit angle of attack (Brooks & Ransbury, 2019). In an unexpected airplane upset event, interrelated factors in aircraft handling, inability to comprehend unfamiliar stimuli, and the psychological stressors of surprise, startle, and fear can combine to create compound threats (Brooks & Ransbury, 2019). Upset recognition and recovery skills require timely and rapid application of corrective inputs, skills that needs to be trained. Upset prevention and recovery training (UPRT) become a focus area of airline operations and training (Rogers et al., 2009). The use of high-end FSTDs for the delivery of UPRT during flight training, complements the application of knowledge and techniques introduced through on-aeroplane UPRT (ICAO, 2014b). The operational potential of FSTD allows for training in upset areas, i.e., low, or very high altitude or in adverse weather conditions, that can be deemed unsafe or impracticable in real aircraft (ICAO, 2015; Miglior, 2014). Growing evidence (Leland et al., 2009; Rogers et al., 2009) suggests that low-cost simulation systems deliver comparable training transfer in UPRT at the fraction of the cost of the FFS. It is essential to establish whether the design of a VR-enabled flight training device will provide effective UPRT training to ensure equivalent safety and effectiveness with the FSTD considering pilot core competencies. This exploratory research addresses two aims. First to determine the effects of the pilot's presence, task-related stress and cybersickness on situational awareness during upset prevention, and, if necessary, recovery. Second to assess pilot acceptance of VR-enabled flight training device in UPRT. The quality of upset recovery training, including the delivery method, the training content, and the training transfer are excluded from this study.

To perform the tasks of upset prevention and recovery, a flight crew needs to deploy several competencies (ICAO, 2014b). Situational awareness and decision-making are critical competencies during the prevention phase, while the application of procedures and aeroplane flight path management - manual control (ICAO, 2014b) are the most critical competencies during recovery from an upset condition. Therefore, the key criterium in the selection of the evaluative scenarios would be an application of situational awareness in prevention and manual control for recovery phases. This competency is well suited to an evaluation in the VR-based environment considering technological limiting factors. Loss of situational awareness among pilots is a well-researched topic (Endsley, 1995; Endsley et al., 2000; Endsley & Jones, 2016; Jones & Endsley, 1996; Stark et al., 2001; van de Merwe et al., 2012). Pilot's errors on the flight deck are typically attributed to disruptions in the decision-making process, however, according to Endsley (1995), it is not the response to the situation but limited or impaired perception and comprehension of the situation - the actual causal factors of accidents. Decision-making relies on situational awareness as a critical factor (Ommerli, 2019), achieving situational awareness is cognitively demanding and it is central to task performance (Endsley, 1995).

One essential part of the UPRT is the skill of recognising all required stimuli and processing the information in upset conditions, and not only from memorised procedures (Brooks & Ransbury, 2019). Pilot situational awareness can be impaired when exposed to physical (i.e., vibration, temperature, lighting, fatigue) or psychological stress (i.e., workload, time pressure, fear, or uncertainty) (Hockey, 1986). There are few documented symptoms of stress factors influencing situational awareness, like a narrow field of attention (focus on a limited number of central cues), cognitive tunnel vision (sampling only obvious or probable sources of information) and premature closure (deciding without exploring all options) (Endsley, 1995). These symptoms affect predominantly the early stage of the decision-making process involving perception, as a result, the assessment of the situation and projection of near-future events (i.e., how the situation may evolve) are impacted by limited recognition of the elements and attributes of a system. Application of situational awareness in prevention and manual control for recovery phases is well suited to evaluate in the VR-based environment considering technological limiting factors.

The impression of being in the virtual environment is a state of psychological awareness commonly referred to as presence (Slater & Wilbur, 1997). In other words, the greater degree of presence, the more likely that humans exposed to VR will perceive the environment and react in a manner like their behaviour in the real world (Slater & Wilbur, 1997). According to Steuer and Reeves (1992) presence is one of the key defining features of virtual environments. The relationship between situational awareness and presence is well documented, and studies (He et al., 2018; Jung et al., n.d.; Prothero et al., 2016) report a significant positive association between these two constructs. Physical and visual motion in a simulator can cause a side-effect known as motion sickness, commonly referred to in this context as simulator sickness, with symptoms including visual disturbances, a decline in hand-eye coordination and gastrointestinal manifestations (Webb, 2010). Additional studies indicate also high severity of those symptoms with delayed effects (Kolasinski &

Gilson, 1998). In the context of virtual reality, simulator sickness is referred to as cyber sickness. Saredakis et al (2020) report that the key impactors on cybersickness are visual stimulation i.e., the content presented in VR, resolution and refresh rate, exposure time and level of locomotion.

Method

Within-subjects nonexperimental design was used. All recruited participants (n = 11) selected for the study were professional pilots with a background in military aviation with a total flying experience mean value of 1377 flight hours (SD = 1286). No power calculations for the sample were conducted due to the exploratory nature of the study.

The simulation software used in this study is based on a commercial off-the-shelf (COTS) X-Plane 11 system. Robin DR401 CDI with Garmin G1000 Electronic Flight Instrument System a general aviation type aeroplane was used during simulated upset scenarios The aircraft selection was a consequence of matching the aircraft as close as possible to the models used in the pilot's training represented by participants' sample. A self-assembled, fixed-base, VR FTD hardware platform was used in the study and consisted of the flight stick and throttle quadrant, Rudder pedals, and VR head-mounted display. HP Reverb G2 VR head-mounted display set has built-in audio capabilities to ensure that the auditory cues were provided during the flight scenarios.

UPRT scenarios emphasize the need for the pilot to maintain situational awareness to recognize a divergence from nominal conditions as early as possible and immediately take corrective action including managing the energy, arresting the flight path divergence, and recovering to a stabilized flight path (ICAO, 2014a). The ICAO (2014a) recommends 16 training that grouped by upset-inducing topics, with each topic consisting of the exercise conditions, training description and rationale. Four selected scenarios for the study (see Table 1) focused on maintaining situational awareness to immediately take corrective action in case of recognised upset by manipulation of the control surfaces and the throttle to maintain the aircraft attitude and correct, if necessary, to return the aircraft to a stabilized flight path.

Table 1: Four selected ICAO flight upset scenarios

Scenario/ Rationale	**Task**
S1. High-altitude upset with environmental factors as a causal factor	Task: Change altitude while maintaining airspeed.
S2. Clean configuration approach-to-stall (high altitude)	Task: Maintain altitude. Reduce thrust to less than adequate. Recognize the stall warning and perform the stall recovery procedure.
S3. Loss of pilot situational awareness leading to LOC-I	Task: Change altitude while maintaining airspeed in IMC.
S4. Energy management leading to performance decrement	Task: Change altitude while accelerating.

Subjective situational awareness was measured immediately after each flight scenario using 10-D Situational Awareness Rating Scale (SART; Taylor, 1990) as derived from a multi-dimensional characterisation of situational awareness consistent with the theory of perception, attention, and cognition (Endsley, 1995; Taylor, 1990). Participants, based on their task performance, subjectively rated each dimension on a seven-point scale. 3-D SART dimensions, attentional demand, attentional supply, and understanding factors were formed from 10-D version of the SART scale (Taylor, 1990). The overall SART score was calculated for each participant for each flight scenario. Pilot rating of task-related stress, as an indirect measure of psychological fidelity used a 24-item Short Stress State Questionnaire (Helton, 2004), as a rating of distress, engagement, and worry states. Stress state was measured pre-test and after each flight scenario and scored on a 5-point Likert scale

(Helton & Nöswall, 2010). In order to analyse the change in the stress profile, the post/pre ratio scores (i.e., differential state changes) (Helton & Nöswall, 2015) were calculated. The concept of presence, a human awareness phenomenon, was operationalised as a measure of attention effectiveness in virtual environments (Witmer & Singer, 1998). The 29-item Presence Questionnaire version 3 (PQ; Witmer et al., 2005) was administered post-test and measured presence through four dimensions (factors): involvement, sensors fidelity, adaptation and immersion and interface quality (Witmer et al., 2005). The overall composite PQ score was calculated as a sum of each factor for each participant. Pilot's wellbeing and acceptance were measured through standardized questionnaires administered post-test. Usability was operationalised as an indirect measure of pilot's acceptance (Burney et al., 2017; Golden et al., 2004; Holden & Rada, 2011). As recommended by Lewis and Sauro (2017), participant's estimation of system level usability was measured using a unidimensional, 10-item SUS questionnaire (Brooke, 1996). Cybersickness was operationalised as an indirect measure of pilot's wellbeing. Simulator Sickness Questionnaire (SSQ) was used to determine the severity of the sickness symptoms induced by the VR simulator (Kennedy et al., 1993) and experienced by participants. Three factors measure, i.e., nausea, disorientation, and oculomotor distress structure, as well as an overall sickness severity were analysed (Kennedy et al., 1993).

Results

This exploratory study examined the effects of pilot psychological experiences, i.e., task-related stress, presence, and cybersickness, on situational awareness during VR exposure during upset prevention and recovery training. SPSS statistical software was used in the data analysis. All results were considered significant at an alpha level $p = 0.05$.

Situational awareness

SART overall score is a function of attentional demand, attentional supply and understanding dimensions (i.e., 3-D), similarly scores for attentional supply and understanding are functions of a wider set of dimensions (10-D).

Overall SART score comparison. As the analysed data failed Mauchly's test of sphericity ($p < 0.05$), to evaluate the effects of the flight upset scenarios on participant's perceived situational awareness, a one-way within-subjects repeated ANOVA with a Greenhouse-Geisser correction was conducted for SART overall score. The different flight scenarios had significant effect on overall SART score, $F(1.88, 18.76) = 6.25$, $p < 0.05$, η^2 (partial) = 0.39, with medium magnitude of the effect (Cohen, 1992). Post-hoc pairwise comparisons with Bonferroni correction showed two significant differences between S1 VFR upset scenario and S3 IFR upset scenario ($M\Delta = 4.64$, $SE = 1.22$, $p = 0.02$) and between S2 stall scenario and S3 IFR upset scenario ($M\Delta = 7.27$, $SE = 1.44$, $p < 0.05$). No significant interactions were found between other flight scenarios.

SART Attentional demand score comparison. The different flight scenarios had significant effect on attentional demand score, $F(3,30) = 14.72$, $p < 0.001$, η^2 (partial) = 0.60, with large magnitude of the effect (Cohen, 1992). Post-hoc pairwise comparisons with Bonferroni correction showed three significant differences between S1 VFR upset scenario and S3 IFR upset scenario ($M\Delta = -7.09$, $SE = 1.02$, $p < 0.001$), between S2 stall scenario and S3 IFR upset scenario ($M\Delta = -7.91$, $SE = 1.52$, $p = 0.002$) and between S3 IFR upset scenario and S4 energy management ($M\Delta = 6.64$, $SE = 1.66$, $p = 0.02$). No significant interactions were found between other flight scenarios.

SART Attentional supply score comparison. The assumption of normality for SART Attentional supply score was verified by inspection of the normal Q-Q Plots and Shapiro-Wilk's test and it was determined to be normally distributed ($p > 0.05$) for most flight scenarios and marginally acceptable for second flight scenario ($p = 0.03$). Attentional supply score between different flight scenarios were not statistically different $F(3,30) = 1.86$, $p = 0.16$, η^2 (partial) = 0.16, with small magnitude of the effect (Cohen, 1992).

SART Understanding score comparison. Understanding score between different flight scenarios were not statistically different $F(3,30) = 2.59$, $p = 0.07$, η^2 (partial) = 0.21 with small magnitude of the effect (Cohen, 1992).

Task-related stress state

An effect significance for each flight scenario was followed up, where applicable, with post-hoc pairwise comparisons with Bonferroni correction to control for the probability of committing a type I error.

Engagement state task comparison. As the collected data failed Mauchly's test of sphericity ($p < 0.05$), to evaluate the effects of the VR exposure on participant's stress state, a one-way within-subjects repeated ANOVA with a Greenhouse-Geisser correction was conducted for SSSQ engagement dimension, for pre-test and each post-task reported stress level. A single factor, the flight scenario, was used during the analysis (i.e., VFR upset vs stall vs IFR upset vs energy management). It was determined that the effect of stress state change in engagement levels were not statistically different between pre-test and post-flight upset scenarios ($F(2.17, 21.86) = 0.81$, $p = 0.47$, η^2 (partial) = 0.08) with negligible magnitude of the effect (Cohen, 1992).

Distress state task comparison. It was determined that the effect of stress state change in distress levels did not elicit a statistically significant change between pre-test and post-S1 VFR upset scenario ($Z = -1.23$, $p = 0.22$), pre-test and post-S2 stall scenario ($Z = -1.08$, $p = 0.28$), pre-test and post-S3 IFR upset scenario ($Z = -0.32$, $p = 0.71$), or pre-test and post-S4 energy management scenario ($Z = -0.37$, $p = 0.72$).

Worry state task comparison. As the collected data failed Mauchly's test of sphericity ($p < 0.5$), to evaluate the effects of the VR exposure on participant's stress state, a one-way within-subjects repeated ANOVA with a Greenhouse-Geisser correction was conducted for SSSQ worry dimension, for pre-test and each post-task reported stress level. A single factor, the flight scenario, was used during the analysis (i.e., VFR upset vs stall vs IFR upset vs energy management). It was determined that the effect of stress state change in worry levels was not statistically different between pre-test and post-flight upset scenarios ($F(1.85, 18.50) = 2.04$, $p = 0.16$, η^2 (partial) = 0.17) with a weak magnitude of the effect (Cohen, 1992).

Presence

Considering the magnitude of scores for each PQ factor, the sensory fidelity ranked the lowest at 58.45% of the maximum score, followed by the involvement factor at 74.2%, and interface quality at 78.80%. Adaptation and immersion PQ factor ranked the highest at 85.5% of the maximum score. The overall results demonstrate a moderate presence score of 74.56%.

Pilot's wellbeing and acceptance

Cybersickness. The descriptive statistics associated with cybersickness as measured post-test using SSQ (Kennedy et al., 1993) are reported in Table 2. Considering the magnitude of scores disorientation ranked the highest most severe factor with the broadest range. The fullness head, blurred vision and dizziness were the symptom profiles elicited by the test conditions that impacted the most. Oculomotor disturbance being the second most severe factor was affected primarily by general discomfort, headache, eye strain and blurred vision symptom profiles. Nausea, the lowest symptomatic factor, was primarily impacted by general discomfort, increased salivation, and sweating. The overall results demonstrate moderate sickness severity, with a relatively broad range.

Threshold values originally proposed by Stanney (1997), would classify the severity as concerning, and extended sickness severity analysis shall be considered.

Table 2: Descriptive statistics for nausea, oculomotor, and disorientation cybersickness factors as measured using SSQ. Note: M = mean; N = frequency; SD = standard deviation.

SSQ Factor group	N	M	SD	Range
Nausea	11	10.47	11.64	0 ÷ 28.62
Oculomotor Disturbance	11	14.47	12.43	0 ÷ 37.90
Disorientation	11	18.98	25.87	0 ÷ 69.60
Overall sickness severity	11	16.32	16.23	0 ÷ 44.88

No participant dropout was recorded during the study. Singular VR exposure mean time was less than 10 minutes for all participants, meaning that each participant was exposed to the VR system for 50 minutes or less for the total duration. The exposure duration was not recorded as a variable but was measured during each test to control for possible exceedances.

System Usability. The SUS results demonstrate a high usability score (M = 86.13, SD = 8.90, Range = 70.00 ÷ 97.50) and it is graded as "excellent" according to SUS adjective rating (Bangor et al., 2008). Systems with scores of less than 72 as marginally acceptable and extended usability analysis shall be considered. Based on the positive score of the investigated immersive flight training system across four different flight upset scenarios, there is an indication that the system has a high acceptance rate among participants in the context of UPRT application.

Relationship analysis

This exploratory research aimed to determine the effects of the pilot's presence, task-related stress and cybersickness on situational awareness observed under four different flying upset scenarios. Correlational analysis was applied to determine the relationship between the variables. A nonparametric measure of Spearman's rank-order correlation was run.

It was determined that the strength of the relationship between situational awareness as measured with SART and self-reported stress state, as measured with SSSQ, was insignificant for the majority of scenarios. However, in-flight scenario three, there was a strong, positive correlation between SART attentional demand factor and SSSQ engagement factor, which was statistically significant ($r_s(9) = 0.61$, $p = 0.048$). Moreover, in-flight scenario four, a strong, negative correlation between SART overall score and SSSQ engagement factor, was also statistically significant ($r_s(9) = -0.75$, $p = 0.008$), as was for SART attentional supply and SSSQ engagement ($r_s(9) = -0.61$, $p = 0.049$), and between SART understanding and SSSQ distress ($r_s(9) = 0.62$, $p = 0.043$). It was determined that the strength of the relationship between situational awareness and self-reported presence, measured with PQ, was insignificant for all scenarios. Similarly, no significant relationship between situational awareness and cybersickness, measured with SSQ, was recorded. Negative correlation between cybersickness and another VR construct – presence, was recorded. The resulting strong negative correlation coefficient, $r_s (9) = -0.62$, was statistically significant, $p = 0.40$. Furthermore, a significant negative relationship was also found between presence and SSQ nausea factor ($r_s (9) = -0.79$, $p < 0.01$) and between presence and SSQ oculomotor disturbance factor ($r_s (9) = -0.63$, $p < 0.05$). The results from correlational analysis indicate potential interactions between the minority of variables. A small sample size should be considered as a partial explanation of lack of statistical significance of relationship between the variables of interest.

Discussion

This exploratory study aimed to establish whether a VR-enabled flight training device will provide effective UPRT training to ensure equivalent safety with the FSTD considering pilot situational awareness competency. It was achieved by determining the effects of the pilot's presence, task-related stress and cybersickness on situational awareness during upset prevention, and, if necessary, recovery and by assessing pilot acceptance of VR-enabled flight training device in UPRT. For this, professional and qualified pilots with sizeable flying experience were selected for the research. The results can be interpreted and implemented in other domains considering certain limitations of the research.

Application of situational awareness in prevention and manual control for recovery phases was the key criterium in the selection of the evaluative scenarios. Considering the overall SART score alone, two significant positive differences in self-reported level of situational awareness between VFR upset scenario (Scenario 1) and IFR upset scenario (Scenario 3) and between the stall scenario (Scenario 2) and IFR upset scenario (Scenario 3) were reported. This is expected and can indicate a lower overall level of situational awareness related to deprivation of visual cues in IMC conditions but increased demand for information exclusively from the instruments to substitute that constraint. Recognition of all required stimuli and processing the information in upset conditions are essential skills acquired by pilots during UPRT. Low attentional demand across VFR scenarios can suggest a low level of perceived variability, and complexity of the situation. High attentional supply across all scenarios indicated pilot's optimal arousal level, spare mental capacity, and good concentration. A moderate level of reported understanding dimension can indicate acceptable quality and quantity of information and general familiarity with the glass cockpit instruments used in the study. Considering perceived situational awareness alone, given the low level of prior exposure to the VR technology, and high proficiency in upset management (as confirmed during flight scenarios), the result of the analysis indicates a lack of negative consequences of VR application on situational awareness needed for effective upset prevention. One limiting factor related to the methodology employed to collect the data, regardless of the validity of the construct, the temporal characteristic of pilot's situational awareness must be considered to establish the full extent of the effect of the VR on situational awareness across the whole duration of the flight. Pilot situational awareness can be impaired when exposed to physical or psychological stress. As these symptoms affect predominantly perception it was difficult to interpret the results due to the adaptation of three factors, engagement, distress, and worry. The result indicates high engagement and low levels of reported distress and worry across all flight scenarios. High attentional demand observed in IMC flight conditions correlated with high engagement may suggest more cognitive resources diverted to perceive the information presented on flight instruments. Lower attentional supply resulted in a higher level of engagement, as observed in scenario four (energy management) would be expected as higher concentration, and attention is required limiting spare mental capacity and engaging the pilot more. Similar observation has been made between understanding factor and distress, as negative correlation demonstrates that the better information and familiarity with the VE, the lower distress. Although SSSQ method is claimed as valid and reliable (Helton, 2004), an objective methodology could be employed to factor in the dynamic character of the scenario and individual characteristics of pilots when performing under stress conditions (i.e., HRV or Galvanic Skin Response meters). It would be expected that stress affects pilot situational awareness during the upset, but the extent of the impact cannot be established. Furthermore, the use of VR in the study additionally complicates the interpretation of results and potentially dilutes content validity.

Presence scores were not found to be significantly correlated with situational awareness in all flight scenarios. High adaptation and immersion PQ factors correlate with high usability scores indicating low entry barriers and overall high engagement in the VR content. A low sensory fidelity score was expected as the interaction with the virtual content, lacks a naturalistic feeling of controlling virtual content and shall be considered primitive. Low sensory fidelity could potentially have an indirect effect on pilot's perception, but at this stage, this argument is not verified, and it is speculative. The results cannot support the argument that the construct of presence is affected by situational factors or vice-versa (He et al., 2018; Jung et al., n.d.; Prothero et al., 2016).

The results of the cybersickness survey demonstrate that the VR-based FTD delivers moderate sickness severity. Although the score of 16.2 is lower than other cybersickness severity scores published in the research literature (Kolasinski & Gilson, 1998; Saredakis et al., 2020; Stanney et al., 1997; Webb, 2010), caution needs to be exercised. The total severity score is close to being classified as concerning, and extended sickness severity analysis shall be considered. The results above 20 would automatically classify the system as a “bad simulator”, but as reported by Stanney (1997) the average sickness severity experienced by participants in virtual environment systems is on average three times higher than when using flight training devices. As expected, (Seay et al., 2002; Weech et al., 2019), a negative significant correlation between cybersickness and presence was recorded. The most probable cause is the diversion of participant attention form from unwanted symptoms, e.g., sensory conflict while experiencing higher presence levels.

The second aim of the study was to assess participant acceptance of VR technology-enabled flight training devices in UPRT. The concept of applying usability measures as a proxy of user acceptance (Burney et al., 2017; Golden et al., 2004; Holden & Rada, 2011) requires understanding of participant behaviour, and attitudes toward technology to guarantee effective, efficient, and satisfactory operation (Holden & Rada, 2011). A high usability score was recorded for the study, and the acceptance of the VR FTD was rated as “excellent” according to SUS adjective rating (Bangor et al., 2008) confirming positive and unsolicited feedback from participants during and after the study.

The key limitation of this study is exclusive employment of a sample with military aviation background with a sizeable flying experience. All participants indicated prior completion of UPRT or comparable training programme with experience in stall conditions and extreme attitudes, exceeding any UPRT requirements for pilots in the commercial aviation sector. While recognising wider upset training syllabubs, the core pilot competencies, as defined by ICAO, must be acknowledged if generalising the study.

The growing evidence suggests that upset-recovery training can be delivered using alternative VR-based TFD (Groen et al., 2012; Leland et al., 2009; Ommerli, 2019; Rogers et al., 2009). The results from this study partially confirm this claim. No evidence has been found that situational awareness was negatively affected by exposure to VR, with certain reactions to stimuli degradation (i.e., flight upset in IMC) resulting in a predictable outcome of increased attentional demand. Limited correlation has been found between situational awareness and task-induced stress, and no relationship was found between situational awareness and presence or cybersickness. The cybersickness severity score is close to being classified as concerning, and extended sickness severity analysis shall be considered. A high participant acceptance score was recorded for the study, and it is graded as “excellent” according to the adjective usability rating.

References

Airbus. (2022). Airbus Statistical Analysis of Commercial Aviation Accidents Statistical Analysis of Commercial Aviation Accidents.

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. International Journal of Human–Computer Interaction , 24(6), 574–594. https://doi.org/10.1080/10447310802205776

BEA. (2012). Final Report on the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro - Paris.

Boeing. (2021). Boeing Statistical Summary of Commercial Jet Airplane Accidents - Worldwide Operations 1959-2020.

Brooke, J. (1996). SUS - a quick and dirty usability scale. In Usability Evaluation in Industry (1st ed.). Taylor & Francis.
Brooks, R., & Ransbury, P. (2019, August 16). Why Upset Prevention Training Alone Is Not Enough. Aviation Performance Solutions. https://apstraining.com/resource/why-upset-prevention-training-alone-is-not-enough/
Burney, S., Ali, S. A., Ejaz, A., & Siddiqui, F. A. (2017). Discovering the Correlation between Technology Acceptance Model and Usability. IJCSNS International Journal of Computer Science and Network Security, 17(11).
Cohen, J. (1992). A Power Primer. Psychological Bulletin [PsycARTICLES, 112, 155.
Endsley, M. (1995). Toward a theory of situation awareness in dynamic systems. Human Factors, 37(1), 32–64. https://doi.org/10.1518/001872095779049543
Endsley, M., & Jones, D. (2016). Designing for Situation Awareness An Approach to User-Centered Design. In Designing for Situation Awareness (2nd ed.). CRC Press. https://doi.org/10.1201/B11371
Endsley, M., Sollenberger, R. L., Nakata, A., & Stein, E. S. (2000). Situation Awareness in Air Traffic Control: Enhanced Displays for Advanced Operations.
Golden, W., Acton, T., Gudea, S., & Scott, M. (2004). Usability and Acceptance in Small-Screen Information Systems. ECollECTeR Conference Proceedings, 6.
Groen, E., Ledegang, W., Field, J., Smaili, H., Roza, M., Fucke, L., Nooij, S., Goman, M., Mayrhofer, M., Zaichik, L., Grigoryev, M., & Biryukov, V. (2012). SUPRA - Enhanced upset recovery simulation. AIAA Modeling and Simulation Technologies Conference 2012, 14. https://doi.org/10.2514/6.2012-4630
He, Z., Zhu, F., Perlin, K., & Ma, X. (2018). Manifest the Invisible: Design for Situational Awareness of Physical Environments in Virtual Reality. https://doi.org/10.48550/arxiv.1809.05837
Helton, W. S. (2004). Validation of a Short Stress State Questionnaire. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48(11), 1238–1242. https://doi.org/10.1177/154193120404801107
Helton, W. S., & Nöswall, K. (2015). Short Stress State Questionnaire: Factor structure and state change assessment. European Journal of Psychological Assessment, 31(1), 20–30. https://doi.org/10.1027/1015-5759/A000200
Hockey, G. (1986). Changes in operator efficiency as a function of environmental stress, fatigue, and circadian rhythms. Handbook of Perception and Human Performance, 2, 44-1-44–49.
Holden, H., & Rada, R. (2011). Understanding the Influence of Perceived Usability and Technology Self-Efficacy on Teachers Technology Acceptance. Journal of Research on Technology in Education, 43(4), 343–367. https://doi.org/10.1080/15391523.2011.10782576
ICAO. (2014a). Doc 10011, Manual on Aeroplane Upset Prevention and Recovery Training. In Doc 10011 AN/506 (Issue 1, p. 102). International Civil Aviation Organization.
ICAO. (2014b). Manual on Aeroplane Upset Prevention and Recovery Training (Doc 10011 AN/506; p. 102). International Civil Aviation Organization.
ICAO. (2015). Doc 9625 Manual of Criteria for the Qualification of Flight Simulation Training Devices - Volume I - Aeroplanes. In DOC 9625-1 (Issue 4, p. 616). International Civil Aviation Organization.
Jones, D., & Endsley, M. (1996). Sources of situation awareness errors in aviation. Aviation, Space, and Environmental Medicine, 67(6), 507–512.
Jung, D., Jo, S., & Myung, R. (n.d.). A Study of Relationships between Situation Awareness and Presence that Affect Performance on a Handheld Game Console. Proceedings of the 2008

International Conference in Advances on Computer Entertainment Technology - ACE '08. https://doi.org/10.1145/1501750
Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. The International Journal of Aviation Psychology, 3(3), 203–220. https://doi.org/10.1207/S15327108IJAP0303_3
Kolasinski, E. M., & Gilson, R. D. (1998). Simulator Sickness and Related Findings in a Virtual Environment. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 42(21), 1511–1515. https://doi.org/10.1177/154193129804202110
Leland, R., Rogers, R. O., Boquet, A., & Glaser, S. (2009). An Experiment to Evaluate Transfer of Upset-Recovery Training Conducted Using Two Different Flight Simulation Devices.
Lewis, J. R., & Sauro, J. (2017). Revisiting the Factor Structure of the System Usability Scale. Journal of User Experience, 12(4), 183–192.
Miglior, M. (2014). UPRT and Simulators (p. 21). Ecair Aviation.
NTSB. (2009). Loss of Control on Approach Colgan Air Inc. Operating as Continental Connection Flight 3407 Bombardier DHC-8-400 N200WQ Clarence Center New York February 12 2009.
Ommerli, C. (2019). Examining the Effects of Perceived Telepresence, Interactivity, and Immersion on Pilot Situation Awareness During a Virtual Reality Flight Exercise [Carleton University]. https://doi.org/10.22215/etd/2020-13923
Prothero, J. D., Hoffman, H. G., Parker, D. E., Furness, T. A., & Wells, M. J. (2016). Foreground/Background Manipulations Affect Presence: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2, 1410–1414. https://doi.org/10.1177/154193129503902111
Richards, N. D., Gandhi, N., & Bateman, A. J. (2012). Improved upset recovery strategies through explicit consideration of pilot dynamic behavior. AIAA Guidance, Navigation, and Control Conference 2012, 15. https://doi.org/10.2514/6.2012-4821
Rogers, R. O., Boquet, A., Howell, C., & Dejohn, C. (2009). An Experiment to Evaluate Transfer of Low-Cost Simulator-Based Upset-Recovery Training. United States. Department of Transportation. Federal Aviation Administration. https://doi.org/10.21949/1503647
Saredakis, D., Szpak, A., Birckhead, B., Keage, H. A. D., Rizzo, A., & Loetscher, T. (2020). Factors Associated With Virtual Reality Sickness in Head-Mounted Displays: A Systematic Review and Meta-Analysis. Frontiers in Human Neuroscience, 14. https://doi.org/10.3389/FNHUM.2020.00096
Seay, A. F., Krum, D. M., Hodges, L., & Ribarsky, W. (2002). Simulator sickness and presence in a high field-of-view virtual environment. Conference on Human Factors in Computing Systems - Proceedings, 784–785. https://doi.org/10.1145/506443.506596
Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. Presence: Teleoperators and Virtual Environments, 6(6), 603–616. https://doi.org/10.1162/PRES.1997.6.6.603
Stanney, K. M., Kennedy, R. S., & Drexler, J. M. (1997). Cybersickness is Not Simulator Sickness. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 41(2), 1138–1142. https://doi.org/10.1177/107118139704100292
Stark, J. M., Comstock, J. R., Prinzel, L. J., Burdette, D. W., & Scerbo, M. W. (2001). A Preliminary Examination of Situation Awareness and Pilot Performance in a Synthetic Vision Environment. Proceedings of the Human Factors and Ergonomics Society, 40–43. https://doi.org/10.1177/154193120104500208

Steuer, J., & Reeves, B. (1992). Defining Virtual Reality: Dimensions Determining Telepresence. Journal of Communication, 42(4), 73–93.
Taylor, R. M. (1990). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In Situational Awareness in Aerospace Operations (pp. 111–127). NATO-AGARD-CP-478.
van de Merwe, K., van Dijk, H., & Zon, R. (2012). Eye Movements as an Indicator of Situation Awareness in a Flight Simulator Experiment. The International Journal of Aviation Psychology, 22(1), 78–95. https://doi.org/10.1080/10508414.2012.635129
Webb, C. (2010). Simulator Sickness in the MH-47G Simulator.
Weech, S., Kenny, S., & Barnett-Cowan, M. (2019). Presence and cybersickness in virtual reality are negatively related: A review. Frontiers in Psychology, 10(FEB), 158. https://doi.org/10.3389/FPSYG.2019.00158/XML/NLM
Witmer, B. G., Jerome, C. J., & Singer, M. J. (2005). The Factor Structure of the Presence Questionnaire. Presence: Teleoperators and Virtual Environments, 14(3), 298–312. https://doi.org/10.1162/105474605323384654
Witmer, B. G., & Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. Presence: Teleoperators and Virtual Environments, 7(3), 225–240. https://doi.org/10.1162/105474698565686

A process to assess the use of human augmentation technologies in defence

Alison Clerici[1]

[1]QinetiQ, UK

SUMMARY

This research evaluated and adapted the use of the Early Human Factors Analysis (EHFA) process to create a methodology for assessing the use of human augmentation technologies in defence.

KEYWORDS

Human augmentation, Defence, Early Human Factors Assessment

Introduction

The aim of this research was to support and bring forward the introduction of human augmentation technologies to defence, by developing a process to help identify how such technologies could be applied to achieve competitive advantage.

The project evaluated, tested and modified the current Early Human Factors Analysis (EHFA) process (as documented in the MoD EHFA Methodology Guide (2016 Issue 1.2)) to create a process that can be used to assess human augmentation technologies in the context of defence. Defining criteria for what 'success' looks like when the modified EHFA is applied to human augmentation technology was critical in the development of the modified process. The success criteria articulate what the EHFA must do to identify potential risks and benefits from the use of human augmentation technology, thereby supporting subsequent decisions on implementation and operationalisation.

Developing the modified EHFA

A set of success criteria was developed to direct the development of the modified EHFA. These criteria acted as requirements for the modified process; for example, the modified EHFA should enable operational benefits of the human augmentation technology in the specific context to be identified. Use of haptic gloves to enhance Explosive Ordinance Device (EOD) training was selected as an example against which to test the current EHFA (MoD, 2016, Issue 1.2), and identify potential modifications to the process. Human Factors Subject Matter Experts (SMEs) and other SMEs from academia were consulted to provide further insights into how the EHFA process needed to be modified in order to satisfy the success criteria. Exploiting these insights, the modified EHFA was then tested against an example Generation After Next (GAN) human augmentation technology; a Cognitive Implant to enhance attention / concentration when undertaking imagery intelligence tasks.

The modified EHFA

The modified EHFA is shown in Figure 1. The macro-stages of the current EHFA process remain, but there are modifications at every stage. Key additions are:

- A human augmentation considerations table – required to support the collation of baseline information on the human augmentation technology and the context / intended use.
- An ‘ethical concern’ column and scale – required to ensure explicit exposure and subsequent consideration of ethics. The scale allows Human Factors Integration (HFI) Risks, Assumptions, Issues, Dependencies and Opportunities (RAIDO) items to be judged for the level of ethical concern that they pose, based on a defined and appropriately tailored scale.
- A decision point at the end of EHFA Stage 3 ‘Assess’ – required to determine if the EFHA indicates that there is adequate justification to continue with implementation of the human augmentation technology, and thus completion of ‘Plan’, and ‘Implement’ stages.
- A new output addressing the success criteria – required to capture information on operational benefits, capability vulnerabilities, capability development activities, trust considerations and any other considerations relating to the specific nature of the human augmentation technology concerned (such as invasiveness and permanence).

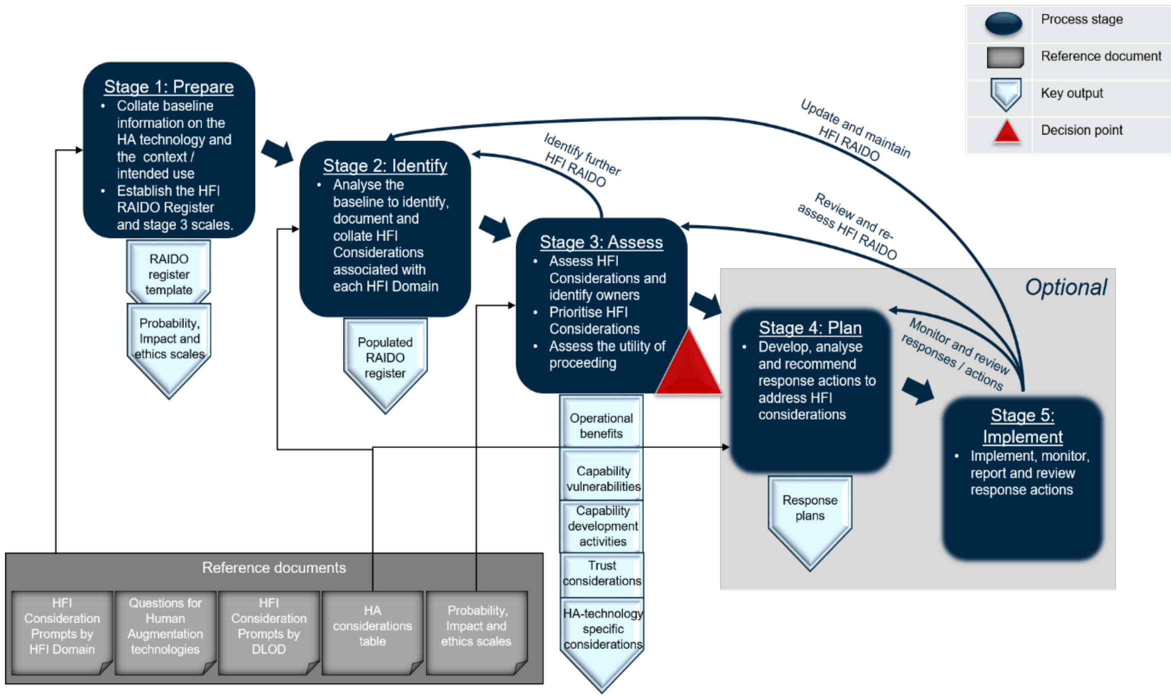


Figure 1: The modified EHFA process, for assessing human augmentation technologies

Conclusions

The modified EHFA process can be used to assess a human augmentation technology for hypothetical implementation into a specific military context. Conducting Stages 1-3 of the modified EHFA will enable decision makers to judge whether a human augmentation technology should be pursued as part of acquisition into a defence or security context. However, further testing and piloting of the modified EHFA process is required in order to refine and validate the process.

Benefits

The modified EHFA process provides a means of establishing the potential benefits of human augmentation technologies, informing and de-risking investment decisions in order to deliver maximum competitive advantage and counter any competitive advantage gained by adversaries exploitation of human augmentation technologies.

References

Ministry of Defence (2016) Early Human Factors Analysis (EHFA) Methodology Guide. Technical Note Issue 1.2

Human Factors Integration Strategy: Embedding Human Factors in Practice within Healthcare

Eva-Maria Carman[1], Giulia Miles[1], Bryn Baxendale[1,2], Emma Smith[2] & Owen Bennett[2]

[1]Trent Simulation and Clinical Skills Centre, Nottingham University Hospitals NHS Trust, United Kingdom
[2]Quality and Patient Safety Team, Nottingham University Hospitals NHS Trust, United Kingdom

SUMMARY

Despite identifying the need for enhancing the use of Human Factors and Ergonomics in healthcare about 20 years ago, progress to date has been slow. A cohesive strategy is required that aligns these methods and expertise with established improvement, transformation and organisational development programmes and which is synergistic with existing work that seeks to address local system and organisational priorities. This paper describes progress to date and proposed future steps for the integration of Human Factors and Ergonomics in one large NHS Hospital Trust.

KEYWORDS

Human Factors Integration, Healthcare, Strategy

Introduction

The need for a greater application of Human Factors and Ergonomics (HFE) in healthcare has been identified as far back as approximately 20 years (Perry et al., 2021), whereby the need for a systems approach to enhance patient safety was highlighted in the seminal reports 'An Organisation with a Memory' (Department of Health, 2000) and 'To Err is Human: Building a Safer Health System' (Kohn et al., 2000). Although these reports have fuelled research interests, the integration of HFE in healthcare has been relatively slow and predominantly focused in certain areas (e.g. patient safety) whereby even in these areas it has been underutilised (Waterson & Catchpole, 2016). This is despite the theoretical models and practical solutions that HFE can offer and the calls by different HFE groups (e.g. CIEHF and Clinical Human Factors Group) for this integration (Waterson & Catchpole, 2016). This highlights that there also needs to be an internal drive within healthcare systems for the integration of HFE. One of the first step towards integration was the UK NHS Concordat on Human Factors and Healthcare (National Quality Board, 2013), signed by 16 healthcare agencies. More recently the establishment of Healthcare Safety Investigation Branch (HSIB, 2023) and embedding HFE principles within the new Patient Safety Incident Response Framework (NHS England, 2022), have introduced new drivers that promote HFE understanding and integration. The aim of this paper is to describe one approach to integrating HFE in a large NHS Hospital Trust.

Context – How the need for an HF Integration Strategy Emerged

The Trent Simulation and Clinical Skills Centre (TSCSC) is a centre for simulation-based education and training within a large acute NHS Hospital Trust. Since opening in 2004, the centre has contributed to enhancing patient care and organisational learning through several programmes of work. The centre's HFE work initially started through supporting several in-house improvement projects, providing training for teams based on the TeamSTEPPS implementation model (AHRQ,

2023) and improving HFE awareness through regular HFE forums for staff. Over time, the HFE team has expanded as has the range and level of involvement on different projects (e.g. service reviews and procurement projects). Due to the larger team, this now includes more in-depth involvement on certain projects, leading the HFE component on multidisciplinary and Trust-wide projects, as well as supervision and mentoring of clinical safety and education fellows. A strategy for HFE is needed to ensure that there is a clear focus for the development of HFE within the organisation and that it is well aligned to the organisation's clinical priorities. The strategy supports the establishment of a HFE team, operating with similar credentials to colleagues in quality improvement and patient safety and underpins the development of the business cases required to attract the necessary resources to build the HFE team.

Strategy – For the Implementation and Integration of HFE across the Trust

The vision for HFE at this Trust is to build on the work achieved to date by a small expert HFE team, broaden the scope of HFE understanding within the Trust and promote the integration and application of HFE to optimise patient safety, staff wellbeing and overall system performance. The strategy has been developed by the HFE specialists within the Trust in partnership with patient safety leads and recognises the need for HFE across numerous workstreams, not just within patient safety work and investigations. This strategy provides a platform on which to introduce HFE integration into organisational functions such as procurement, information technology, estates and facilities, human resources, and similarly into service performance of clinical divisions and integrated care pathway design.

The strategy consists of four guiding principles, namely co-production and co-working, alignment with current organisational functions and workstreams, the expansion of a core team of experienced and qualified HFE specialists and building better HFE capability and capacity within divisions and corporate functions. It recognises the importance of offering internal users a coherent approach to designing services and aligns HFE with patient safety, quality improvement and organisational development, offering 'internal clients' a joined-up and cohesive approach to programmes of work.

In its current form, it describes a preliminary route for the development of HFE capacity and capability through the application of HFE expertise within prioritised projects, enhancing the awareness of HFE with organisational and service leads, and providing access to HFE training for the Trust. The strategy emphasizes the systems perspective, user centred approach and embedding and improving systems and processes by understanding human capabilities, adaptations and "work-as-done". HFE specialists would provide methods, objectivity and solution development for service transformation and safety improvement work as well as provide support and mentoring of staff embedded within clinical and organisational functions, thus growing a cadre of clinical and non-clinical staff with core HFE skills. Elements of this model are already being applied in certain areas, with further development taking place with clinical and corporate divisions considering both local and wider Trust goals and aims to mirror similar integration models in other industries.

Discussion, Next Steps and Conclusion

This strategy for HFE integration proposes to build on how HFE experts have worked in this Trust in the past and expand the HFE offer to the organisation, improving equity of access to HFE advice. It aims to focus both at local clinical service goals while still maintaining sight of organisational priorities and larger work programmes that would benefit from HFE input, as suggested by Perry and colleagues (2021). To date, this strategy has been presented at a variety of different leadership and clinical service meetings within the organisation and feedback is being actively sought from clinical teams about how they would want HFE to be integrated. HFE has been included in the organisation's newly published strategy and next immediate steps include describing in detail how the strategy can be implemented in practice along with identification of the resources needed by outlining business cases, for presentation to the Trust's senior management group. At a time of significant and prolonged stress on the healthcare service, HFE must provide compelling evidence and relevant examples, so the benefits of the HFE strategy are clear to stakeholders and outcomes are aligned and connect with Trust and system wide priorities.

References

AHRQ. (2023). Agency for Healthcare Research and Quality - TeamSTEPPS®. https://www.ahrq.gov/teamstepps/index.html

Department of Health. (2000). An organisation with a memory: Report of an expert group on learning from adverse events in the NHS.

HSIB. (2023). Healthcare Safety Investigation Branch. https://www.hsib.org.uk/

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (2000). To Err Is Human: Building a Safer Health System. National Academy Press.

National Quality Board. (2013). Human Factors in Healthcare: a Concordat. http://bit.ly/1hIsXel

NHS England. (2022). Patient Safety Incident Response Framework. https://www.england.nhs.uk/patient-safety/incident-response-framework/

Perry, S. J., Catchpole, K., Rivera, A. J., Henrickson Parker, S., & Gosbee, J. (2021). 'Strangers in a strange land': Understanding professional challenges for human factors/ergonomics and healthcare. Applied Ergonomics, 94, 103040. https://doi.org/10.1016/J.APERGO.2019.103040

Waterson, P., & Catchpole, K. (2016). Human factors in healthcare: welcome progress, but still scratching the surface. BMJ Quality & Safety, 25(7), 480–484. https://doi.org/10.1136/bmjqs-2015-005074

Human Factors and Procurement: Lessons Learnt from a High-Value Procurement Exercise

Eva-Maria Carman[1], Michael Johnson[1] & Giulia Miles[1]

[1]Trent Simulation and Clinical Skills Centre, Nottingham University Hospitals NHS Trust, UK

SUMMARY

Human Factors can inform and enhance traditional procurement processes by capturing the users' input and considering the wider system into which the products will be implemented. Despite this, traditional procurement processes do not typically consider integrating HFE into the process in a systematic way. This paper describes the role of Human Factors in the different phases of a large procurement project and the lessons learnt for the procurement of hospital beds within one large NHS Hospital Trust. The aim of the procurement project was to determine the best solution that includes a variety of products and service contract from one supplier. A total of six different bedframes and two different types of mattress needed to be considered. The role of the HFE team was to provide advice at strategic project meetings, support the specification design, conduct an HFE evaluation of the products and ensuring a system's perspective was considered throughout the process. Across all the product types, the HFE evaluation included 27 simulation testing sessions followed by feedback from staff on in-situ use on 23 different wards. Key lessons learnt included the value of qualitative data can add to support the decision-making process in procurement projects, the need to understand clinical needs as in this context there is no one perfect product due to the wide range of applications, and the need for HFE specialists to have a better understanding of the procurement process and their involvement across all phases of this type of project.

KEYWORDS

Healthcare, Procurement, Systems Approach

Introduction

As a design science and with users and people at its centre, Human Factors and Ergonomics (HFE) is well placed to support the procurement process. Despite this though, traditional procurement processes do not typically consider integrating HFE into the process in a systematic way, and rather just add it as an afterthought or supplement (Cassano-Piché et al., 2010). Extensive work has been done describing the evaluation of medical products with regards to the general usability at a more micro-ergonomic level, for example the anthropometric and biomechanical aspects (e.g. Adeodu et al., 2014; Kim et al., 2009; Mehta et al., 2011). However, as these products are brought into existing work systems, there is also a need to assess the usability relative to the work system they will be integrated into, which may impact the decisions made during the procurement process.

Human factors-informed procurement enhances the traditional procurement process by including a multidisciplinary team, frontline staff or user participation and HFE expertise to lead on the HFE methods (Cassano-Piché et al., 2010). One key contribution of HFE to the procurement process is the information on usability it can provide, namely the functionality, interface design, and training and learning requirements (Carayon, 2011). In addition to this, HFE can provide the tools to consider the product relative to the users, tasks and environments, which is essential (Wilson & Sharples, 2015). However, to be successfully integrated into the procurement process, one needs to understand the complexities, limitations and structure the process may impose due to legal requirements, as highlighted by Hignett and Lang (2013). The tender process, which may be part of the procurement process, may have the following phases: advertisement, registration of interest, notification of tender specifications, product evaluations, awarding and implementation of the contract. To reap all the benefits HFE can offer procurement projects and to ensure HFE can be appropriately applied, this will require HFE to be integrated early on and throughout the process and not just in the evaluation stage. This paper describes the role of HFE in the procurement process and the lessons learnt for the procurement of hospital beds within one large NHS Hospital Trust.

Method

This procurement project aimed to select a supplier that would provide beds, mattresses, and a service contract for these products at a large NHS Hospital trust that offers a wide range of general services. To support this project and the decision-making process, a diverse project team (i.e. Finance, Procurement, Medical Engineering and Physics, Human Factors, Infection control, Estates, Tissue Viability) that was representative of the key services for maintaining and supplying beds for the service was assembled. The team was responsible for evaluating the service package and different models in the product categories to determine the most suitable ones for a range of different bed types and mattresses for the Hospital Trust. A total of six different bed types (i.e. medical and surgical beds, critical care beds, birthing beds, bariatric beds, paediatric beds and cots, low bedframes) and two different types of mattress (i.e. dynamic surface, foam surface) needed to be considered. The components of the different phases in this procurement project and the role of the HFE team has been depicted in Figure 1.

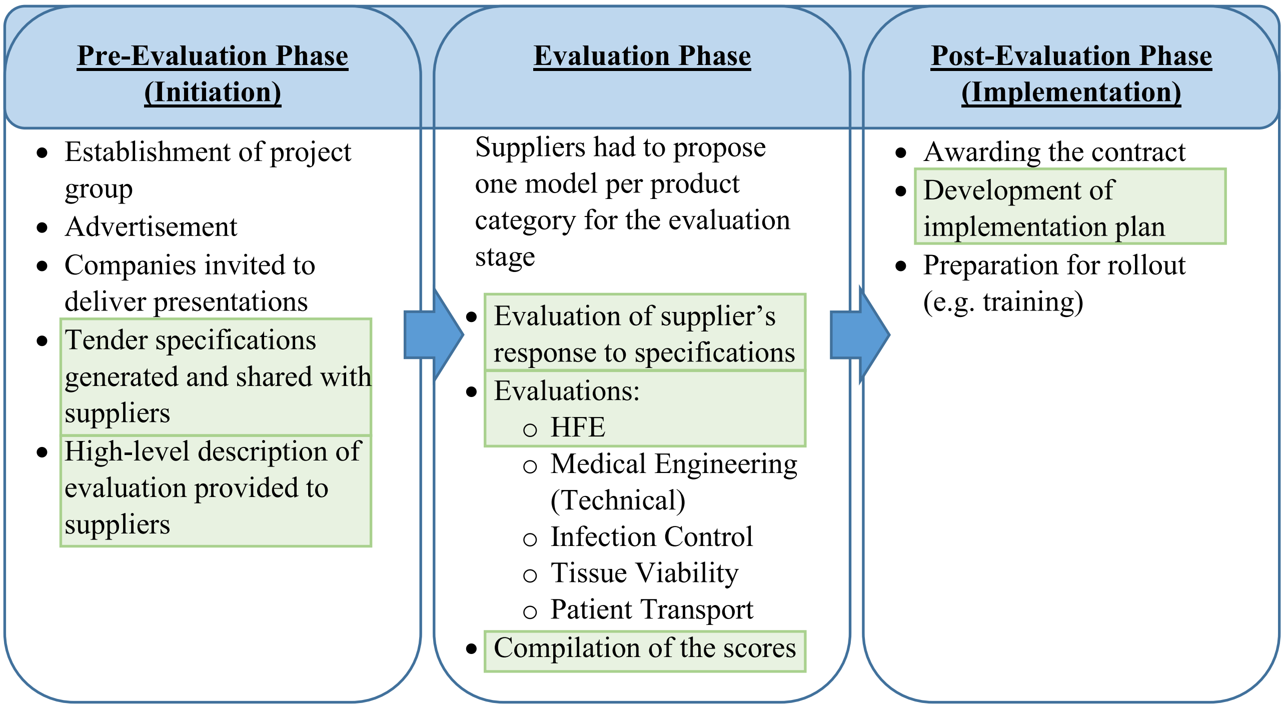


Figure 1: The components of the different phases in this procurement project and the role and involvement of the HFE team (marked in green).

The HFE evaluations focused on the functionality, interface design elements and capturing the user perspective on these products, with particular focus on the work system wherein these products would be used, rather than the micro-ergonomic aspects. The HFE evaluation consisted of two phases at two different time points, simulation testing in a simulation centre followed by feedback from staff on in-situ use (ward testing). These protocols were based on the seven principles of universal design (Story, 1998) and usability heuristics (Nielsen, 1994). Data for both phases was captured using a structured questionnaire that consisted of both open-ended and Likert Scale questions.

The simulation session provided an opportunity to conduct usability walkthroughs and usability tests. Participants were recruited from across the clinical workforce for the products that may be selected for their work area. Frontline staff were provided with an opportunity during the simulation to interact with the products as well as test several clinical scenarios (i.e. CPR, patient transfer and hoist tasks) and then debriefed as a group to collect the group's feedback on the products. Staff were also asked to complete a questionnaire on the products following the simulation session, which included questions on the structural and functional features of the product, comparison to the products currently in use, movement considerations, control panel and display use, evaluation of clinical scenarios as well as the general impression and patient considerations. The open questions asked staff to consider the products compatibility with other pieces of equipment (e.g. oxygen, IV stands), essential features for the patient group, staff requirements for use of the product, training requirements and provided an opportunity for any additional comments. The simulation testing also used group consensus to determine how the product compared to the current model in use, the top three features, worst three features and if any "dealbreakers" were present. The ward questionnaire included questions on complexity regarding the use of the product, function integration, and comparison to the other products currently in use. The open questions asked staff to consider the products compatibility with other pieces of equipment, problems encountered during use, entrapment concerns, top three features, worst three features and if any "dealbreakers" were identified. The questionnaire used for the ward testing also included an adapted System Usability Scale (Brooke, 1996). In both the simulation and ward questionnaires, each Likert-based question was normalized by determining the average score per question across the testing session for all participants (not including the System Usability Scale). Both the ward and simulation testing were weighted equally and percentages were determined for each product for each phase.

The HFE qualitative and quantitative results were fed-back to the project group to support the discussion of the results from evaluations conducted by the other teams (e.g. medical engineering) and support the decision-making process. As the project group required predominantly quantitative results, the qualitative HFE results from both simulation and ward phases were assessed by a panel that included HFE specialists and clinical educators. Each product was given a grade label by the panel based on the summary of key qualitative results from the group discussion in the simulation testing and the open-ended questions from the questionnaires used in both testing phases. The grading scale was defined by the project team and used by the other evaluation teams in earlier phases of the project. The lowest grade (Grade 0) was deemed unacceptable and defined as the product completely failing to meet the required standard or does not provide an answer. The highest grade (Grade 4) was deemed excellent and defined as the product meeting the required standard in all material respects and exceeds some or all the major requirements. Any products that scored a Grade 1 (weak) or 0 (unacceptable) were highlighted to the project team, with specific justification for these grades. The final summary reports submitted to the project group included not only the quantitative scores from the questionnaires but key qualitative results that staff felt were "dealbreakers" as well as highlighting any results that were contradictory to the specifications provided to the suppliers in the pre-evaluation phase. The HFE team were also involved in the discussion of the results, to highlight or draw out any essential results from the HFE evaluation needed for the discussion and to support the development of potential resolutions. Similarly, the results from the other evaluations (e.g. technical, tissue viability) were brought back to the project group and discussed.

Findings: HFE Involvement, Challenges and Lessons Learnt

The aim of this paper was to describe the role of HFE in this type of procurement project, the challenges faced, and the lessons learnt. Several unique characteristics of this type of project naturally resulted in constraints for the project team and the evaluations and as a result created challenges. These characteristics included the time span of these types of projects and the prescribed process by NHS procurement. This project has spanned several years, with some preliminary assessments occurring in 2019 and then the process was restarted in 2020. The simulation testing occurred between September and October 2020 with 27 testing sessions being conducted, one per product model. The participant groups ranged from two to eight participants per session who were clinical staff that would be using these products if the contract was awarded to that supplier. The ward testing was conducted between March and April 2021, and 21 products were placed onto 23 wards for approximately two weeks. Not all products had a patient use them in this time. Staff were asked to complete a questionnaire during this time period and a range of one to 22 questionnaires were returned per ward. An additional risk validation study of certain products was conducted by the HFE team in October 2021 and September to November 2022. Despite this length of time, the project team was very lucky in that all of the core members remained constant and a continuity could be provided throughout the evaluation phase. Another key characteristic of this project was that although the HFE team was called in relatively early in the project timeline, the tender process was already prescribed by NHS procurement. This created some specific requirements for the testing. Due to the prescribed process, the specifications were required early in the process (Figure 1, pre-evaluation phase) and could not be informed by the simulation testing. This highlighted that at the start of the process, the level of detail required in the specifications was not fully known. Another key influence on this project was that of the COVID-19 pandemic. Although this had a negative effect on recruiting participants for the simulation testing, it had a positive effect on the team resources required for the simulation testing. As a result of the pandemic, the simulation centre did not have its usual courses which allowed for space and staff to be available to conduct the simulation testing.

Additional challenges included the resource intensive nature of this project, low overall number of volunteers and low numbers in certain staff groups, the assumption that the “best” products would be put forward by the suppliers, and managing the expectations of staff who participated in the testing phases. This project was resource intensive on multiple levels including the high volume of testing required with limited resources, the team requirements on simulation testing days for the HFE evaluation, and the data analysis which included qualitative analysis. An underlying assumption in this project was that the “best” products on the market would be put forward by the suppliers, however when it came to product testing, frontline staff identified elements that due to various reasons made the product not suitable or the best option for that patient group. This highlights how trade-offs are made and how these products often need to serve a varied patient group and therefore will not be the “best” solution for all. As this project required the team to select a supplier that offered the best combined offer, the expectations of staff involved in the testing phases had to be managed as their preferred product may not have been selected. Furthermore the simulation testing highlighted some of the limitations of the specification design and the lack of range of products for certain categories. These included that the specifications were too vague in certain areas.

Key lessons learnt included the positive effect and benefit of having a multidisciplinary team that supported the HFE evaluations, the value of qualitative data in these types of projects and the need for HFE specialists to have a better understanding of the procurement process. The multidisciplinary HFE team that conducted the HFE evaluations included HFE specialists, technical leads, and multi-professional clinical educators that had a range of clinical backgrounds. This not only allowed the testing to run in a smooth and efficient manner due to the team size and specialty range but also assisted in troubleshooting problems that arose on testing days. The value of the qualitative data was highlighted during the discussion of the products, not only in project team meetings but throughout the process as queries on the products arose. The qualitative data provided tangible examples of usability concerns and context for the quantitative results. Another key lesson learnt was that a better understanding is needed of the requirements and constraints at a project initiation level of the procurement process to design better evaluations and scoring systems. Despite numerous aspects of this project prescribed by NHS Procurement, there was still a degree of evolution in this project. With a better understanding of the inherent constraints and natural evolution of these types of projects by the broader project and evaluation team including the HFE specialists, better interventions that work within these limits could be designed, unsuitable products may be excluded and the limited resources can be utilized to evaluate the key contender products better. This is particularly important as testing needs to be fair and equal but also recognize the resources that may be required, especially in healthcare, may not be available. Additional key lessons learnt included that there are numerous opportunities to potentially enhance design and work with the manufacturers to ensure a "good fit", the need to understand clinical needs as in this context there is no one perfect product due to the wide range of applications required of the products, and the potential role of HFE in the different procurement phases (e.g. initiation, evaluation, implementation). Furthermore, clarity of the HFE role is essential at all stages – HFE Specialists provide an objective, evidence-based approach to support the organisation to make informed decisions about "best fit package" and not just individual product assessments.

Although the focus of this paper was not the results from the HFE evaluation, several of the design and usability concerns were identified that highlighted key lessons learnt. Some of the key design concerns raised through the HFE assessment across the products included complicated staff interface buttons, the design of the head-end in certain models restricted the access to the patient's head, and the visibility and ease of use of the CPR mechanism. Design features identified by staff as limitations that were as a result of trade-offs included "how low was low enough" for beds and the use of adult beds for specialty patient groups (e.g. paediatric patients). Beds that were able to go as low as possible often had other limitations due to the structural requirements needed to allow the lowering of the bed. Several examples were identified whereby specialty patient groups required an alternative product (e.g. paediatric patients requiring general adult medical beds) whereby then certain structural elements were not designed for this population group (e.g. cot sides). These two examples highlighted the compromise the project team needed to find as the patient group to use the product was so broad. This required extensive discussions on the results, additional risk evaluations and additional discussions with frontline staff to determine which elements were critical.

Discussion

A key challenge and aim of the procurement process is to find a balance between the hospital's requirements from the product which may include being used for a range of patient groups, and the terms of the service contract at the best possible financial price (Hignett & Lang, 2013; Western Canada Human Factors Collaborative, 2017). The HFE evaluation in this procurement project aimed to capture elements of usability, frontline's staff perspective of the products and utility with regards to how well these products would fit into the existing work systems. Accordingly, the HFE indicators selected were related to general usability but also wider work system considerations, for example patient consideration, other tool and technology interactions, and different tasks associated with these products.

As highlighted in the document "Guidance for Human Factors Evaluations in the Procurement of Medical Devices, Equipment and Technology" produced by the Western Canada Human Factors Collaborative (2017), four recommended evaluation methods have been proposed for HFE evaluations in procurement activities. These consist of usability walkthroughs, heuristic evaluations, usability testing and field studies. This procurement project utilized three of these four methods namely usability walkthroughs, usability testing and field studies. The benefit of using these three different recommended evaluation methods is that a range of sensitivities, objectivity, control and realism was covered (low, medium and high). Furthermore, the usability walkthrough and usability testing both used task scenarios, which allowed several key design concerns to be identified (e.g. limitations associated with CPR mechanisms). In an earlier phase of this procurement project (April 2019), a heuristic evaluation was included but once the project was restarted in 2020 this was removed for the evaluation design. This was excluded due to the volume of products that needed to be tested in the revised project in 2020 (27 models tested) and the available resources. Furthermore, it was decided to prioritise end-user involvement and the selected methods have better control for unknown variables and improved objectivity (Western Canada Human Factors Collaborative, 2017).

Some of the limitations of this project included that a heuristic evaluation of the products could not be performed, and the limited sample sizes for certain testing sessions for both the ward and simulation testing sessions. Some of the limitations that resulted and challenges experienced by the HFE team were as a result of the procurement process and potentially the limited involvement in the initiation phase. This point has also been highlighted by Hignett and Lang (2013), specifically the limitations associated with the specifications such as insufficient detail and ambiguous results on these in the returned documentation from the suppliers. Greater HFE involvement upfront, particularly in the initiation phase, could assist with the development of more precise specifications, especially defining more specific HFE requirements more clearly.

Key strengths of this project included incorporating a system's perspective throughout this project, the commitment by the organisation to have HFE included in this project, increasing the involvement of end-users and due to consistent results from multiple rounds of testing for certain product groups, the method and tools proved reliable. A system's perspective was brought into this piece of work not only through the HFE evaluations, but by including the HFE team in the project team and in the discussion of all the evaluation results. This also ensured that there was a representative to feedback the opinion of end-users on those products at multiple stages. Furthermore, the strong support from the project manager and the collaborative nature across the different evaluation team members allowed for the resolution of operational issues, allowed for through evaluations to take place and also ensured that the HFE team was supported in feeding back frontline staff's concerns. The co-location of certain stages of the evaluation allowed for issues identified to be checked by the other evaluation teams in an efficient manner.

Conclusion

This paper describes the role of HFE in the procurement process and the lessons learnt for the procurement of hospital beds within one large NHS Hospital Trust. The aim of this procurement project was not to find the best products on the market for certain product types, but rather determine the best solution that includes products and service contract from one supplier that best meets the needs for one Hospital Trust. A key element and contribution of the HFE team to this procurement project was providing the perspective of how these products may or may not fit into the existing work system and therefore had a macro-ergonomic approach. The HFE team was involved to different degrees in the initiation, evaluation and post-evaluation phases and played an active role in the project team. In addition to designing and conducting the HFE evaluation component of the evaluation phase, the team provided strategic advice at project meetings across all three phases, ensuring that a system's perspective and the perspective of frontline staff was considered. This highlights some of the benefits and support HFE can provide the procurement process, namely a system's approach, the user's perspective and objective and varied tools and approaches. Challenges emerged as a result of the procurement process and included the specific requirements and constraints placed on testing due legal requirements, the assumption that the "best" products would be put forward by the suppliers, and managing the expectations of staff who participated in the testing phases. Key lessons learnt included the value of qualitative data can add to the discussion and support the decision-making process, that there are numerous opportunities to potentially enhance design and work with the manufacturers to ensure a "good fit", the need to understand clinical needs as in this context there is no one perfect product due to the wide range of applications, the need for HFE specialists to have a better understanding of the procurement process and their involvement across all phases of this type of project. With a better understanding of the inherent constraints of these types of projects by HFE specialists, better evaluation designs that work within these limits could be created and the limited resources, particularly in healthcare, can be utilized more effectively. As this project is still ongoing and currently in the post-evaluation phase (implementation), there is still a role for HFE as now changes to the work system start to emerge as these products become integrated into numerous subsystems within the existing work system (e.g. device library, complaints system).

Acknowledgements

The authors are grateful to all the colleagues from the Trent Simulation and Clinical skills team as well as the project partners from Nottingham University Hospitals NHS Trust for their support and work on this project. Thank you also to all the frontline staff who gave of their time to participate and evaluate these products both in the ward and simulation testing phases.

References

Adeodu, A. O., Daniyan, I. A., & Adaramola, O. O. (2014). Anthropometry Evaluation Using Analytical Hierarchical Process and Independent Test Method for Hospital Bed Design: A Case Study in Nigeria. Journal of Computation in Biosciences and Engineering. http://scienceq.org/Journals/JCLS.php

Brooke, J. (1996). SUS: a "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. Weerdmeester, & I. L. McClelland (Eds.), Usability Evaluation in Industry (pp. 189–194). Taylor and Francis.

Carayon, P. (2011). Handbook of Human Factors and Ergonomics in Health Care and Patient Safety, Second Edition. Taylor & Francis. https://books.google.co.uk/books?id=7IJMP6UlHUIC

Cassano-Piché, A., Cafazzo, J. A., Chagpar, A., & Easty, A. C. (2010). How Human Factors Methods Can Help in the Procurement Process. Human Factors Horizons, 49–56.

Hignett, S., & Lang, R. (2013). Project-managing a change to electric hospital beds. British Journal of Healthcare Management, 9(8), 271–276. https://doi.org/10.12968/BJHC.2003.9.8.18852

Kim, S., Barker, L. M., Jia, B., Agnew, M. J., & Nussbaum, M. A. (2009). Effects of two hospital bed design features on physical demands and usability during brake engagement and patient transportation: A repeated measures experimental study. International Journal of Nursing Studies, 46(3), 317–325. https://doi.org/10.1016/J.IJNURSTU.2008.10.005

Mehta, R. K., Horton, L. M., Agnew, M. J., & Nussbaum, M. A. (2011). Ergonomic evaluation of hospital bed design features during patient handling tasks. International Journal of Industrial Ergonomics, 41(6), 647–652. https://doi.org/10.1016/J.ERGON.2011.07.005

Nielsen, J. (1994). Heuristic Evaluation: Ten Usability Heuristics. https://pdfs.semanticscholar.org/5f03/b251093aee730ab9772db2e1a8a7eb8522cb.pdf.

Story, M. F. (1998). Maximizing Usability: The Principles of Universal Design. Assistive Technology, 10(1), 4–12. https://doi.org/10.1080/10400435.1998.10131955

Western Canada Human Factors Collaborative, . (2017). Guidance for Human Factors Evaluations in the Procurement of Medical Devices, Equipment and Technology.
Wilson, J. R., & Sharples, S. (2015). Evaluation of Human Work. CRC Press. https://books.google.co.uk/books?id=uXB3CAAAQBAJ

Author Index

Chartered Institute
of Ergonomics
& Human Factors