Ubiquitous Technologies for Capture of Real-World Performance
Sarah Sharples, Michael Brown, James Pinchin, Jesse Blum, Arzoo Nagiyev, Brendan Ryan, Dominick Shaw & John Blakey

Human Factors Research Group, Faculty of Engineering, University of Nottingham, Nottingham, UK; Horizon Digital Economy Research, University of Nottingham, Nottingham, UK; Nottingham University Hospitals NHS Trust, Nottingham, UK; Liverpool School of Tropical Medicine, Liverpool, UK.

Much theoretical understanding of specific relationships between phenomena such as workload, performance and attention in Ergonomics and Human Factors (E/HF) has been based on studies conducted in the laboratory context. This setting is often chosen due to the ability to control variables and apply multiple, often intrusive measures, to capture detailed measures of individual response. However, there has also been considerable emphasis on the importance of capturing data from the ‘real world’ context (see Wilson & Sharples, 2015) and the discipline of E/HF has for many years considered both laboratory and real world approaches. The increasing capacity and decreasing intrusiveness of equipment-based measures, such as physiological measures, location tracking and video cameras, means that the distinction between laboratory and real-world approaches is no longer as distinct as it once was, and it is becoming increasingly possible to apply detailed and previously intrusive measures in a real-world setting, even on occasions when that setting is safety-critical. This paper uses two case studies from the domains of healthcare and transport to examine the practical and theoretical implications and opportunities of applying technologies to capture real-world performance. We focus on the value of these methods for both practical support of ergonomics interventions in workplace design, and also the development of theories to support E/HF knowledge. We also raise important questions about the way in which such technologies can and should be implemented in a workplace context, including ethical considerations, privacy concerns and the ‘use and abuse’ of data that describes workplace performance in a high level of detail, on an individual and organisational level.

Practitioner Summary: Ubiquitous and embedded technologies have the potential to track tasks, movements and behaviours in the workplace. These technologies can enhance the level of detail of data that is collected in a real world context, and ultimately support more effective Ergonomics/Human Factors design and analysis. However, it is important to develop appropriate protocols to deploy such technologies in real world contexts, to ensure that the data collected are valid, effective, ethical and useful.

Keywords: Cognitive ergonomics, performance, healthcare, transport, measuring performance, real-world methods

1. Introduction

Much theoretical understanding of specific relationships between phenomena, such as workload, performance and attention in Ergonomics and Human Factors (E/HF), from early examinations of dual task processing (e.g. Wickens, 1981) to developing understanding of automation complacency has been based on studies conducted in the laboratory context (e.g. Kidwell et al., 2014). The laboratory is the traditional setting of studies where a stimulus or task is manipulated in a controlled manner by the experimenter to produce a set of independent variables; the effect of this manipulation of independent variables on selected dependent variables is then captured. The laboratory setting is often chosen due to the ability to control variables and apply multiple, often intrusive measures, to capture detailed measures of individual response. The experimental paradigm that is then imposed on this setting allows the application of statistical analyses, which can then determine the likelihood of occurrence of the different effects under consideration, compared with a chance effect.

However, there has also been considerable emphasis on the importance of capturing data from the ‘real world’ context (see Wilson & Sharples, 2015) and the discipline of E/HF has for many years considered both laboratory and real world approaches. Moray (1994) and Waterson (2009) have highlighted the importance
of capturing behaviour in context, and there is clear value in understanding the effect of different elements of workplace design on the response of individuals and teams. The capture of real world performance and behaviour is inevitably detailed, time consuming and provides less opportunity for capture of the effects of variables on performance and behaviour.

Some methods do allow the imposition of the experimental paradigm on a real world setting. For example, Sharples et al. (2011) used a structured observation approach to identify the difference in behaviours of rail signallers with different levels of automation. Some methods that have been used in laboratory settings to collect detailed performance data, such as physiological methods or eye tracking, have until recently been considered intrusive and particularly susceptible to the ‘Hawthorne effect’, or the act of observation altering participant response. However, the increasing capacity and decreasing intrusiveness of equipment-based measures, such as physiological measures, location tracking and video cameras, means that the distinction between laboratory and real-world approaches is no longer as wide as it once was, and it is becoming increasingly possible to apply detailed and previously intrusive measures in a real-world setting, even on occasions when that setting is safety-critical.

This paper considers a series of case studies to examine the practical and theoretical implications and opportunities of applying technologies to capture real-world performance.

2. Case Studies

Two case studies are selected to demonstrate the potential for ubiquitous technologies for performance capture. The first is taken from a research programme that examined the impact of automation and novel signalling technologies in the rail industry. The second is taken from a programme of work that is examining the way in which clinicians perform out of hours care in hospitals. Both case studies represent situations which are safety critical, and therefore where it is important that any measures that are applied in the real world do not disturb the participant from completing the main job of work. They also represent contexts where work design changes are being implemented, and where there is therefore value in understanding the impact, or potential impact, on behaviour and performance as a result of these changes: in rail, new technology has the potential to change driving strategies and performance; in the clinical context, rota design, task prioritisation training, technology support for task allocation and ward design could all affect clinical performance. Therefore there is clear value in collecting real world data to support cases for workplace design change, or provide benchmarking data to allow analysis of later workplace alterations.

2.1 Impact of automation in train driving

The first case study considered the impact of the introduction of ERTMS (European Rail Traffic Management System) on the task of train driving. Earlier work in the research programme (Buksh et al., 2013a, 2013b; Naghiyev et al., 2014a, 2014b) had involved conducting interviews and expert observations of driving to understand the way in which driving strategies are likely to change as a result of changes to the way that signalling systems are displayed to train drivers. In the UK rail system, current systems work on a ‘fixed block’ principle where signals are at fixed points at the side of the railway track, and drivers are required to observe and respond to the status of signals whilst driving. ERTMS introduces in cab signalling, and moves away from the fixed block principle (potentially increasing capacity on the railway) as well as presenting an interface to train drivers that suggests a range of speeds that should be targeted whilst driving. In addition to this change to the form of information presented, drivers are also presented with additional displays that show details about the expected times of movement along the track (and movement authorities) and additional automated systems are introduced, for example to sound auditory alarms and provide visual alerts in the case of overspeeding (Naghiyev et al., 2014b).

Whilst the earlier stages of the research programme gave clear indications of the different types of strategies that might be adopted by train drivers as a result of these changes, it was felt to be important to follow up this qualitative data with more quantitative and detailed analyses of the specific ways in which drivers behaved in different parts of a driving journey, such as travelling past level crossings, or approaching station stops. In particular, a research question was established to explore whether the ways in which drivers allocated their visual attention were affected by the change in signalling system.

Therefore eye tracking data were collected from 28 train drivers, 14 of whom were driving using the conventional, fixed block, external signalling system, and 14 of whom were using the new ERTMS system.
(Naghiyev et al., 2014b). In addition, verbal protocol analysis was used to ask drivers to describe their activities and strategies during the train driving journey.

Figure 1 shows an example of the way in which the visual field of the drivers was divided into areas of interest as a basis to determine the percentage of time during a journey or task that the drivers spent looking in a specific area of the visual field. These ‘areas of interest’ were used as as the basis for analysis of fixation durations in different parts of the driver cab, allowing data to be obtained comparing the mean fixation durations inside and outside the cab. Detailed data reported in Nagiyev et al. (2014b) confirmed that, for example, a consequence of the introduction of the new ERTMS display was that drivers had higher average total fixation durations for the speed area of the interface. It also demonstrated that there was high inter-individual variability in strategies adopted by drivers when presented with these new interfaces – a result that has important consequences for driver training and assessment.

![Figure 1. Example areas of interest used as basis for eye-tracking analysis for train drivers.](image)

2.2 Analysis of clinical out of hours work

The second example of application of ubiquitous technologies to analysis of real world performance is taken from the healthcare sector. 75% of hospital care in the UK is ‘out of hours’ (outside the core working hours of 9am-5pm, Monday to Friday) where a small number of clinical staff are responsible for a large number of physically distributed wards around the hospital (Brown et al., 2015). A series of research activities led by the University of Nottingham, in conjunction with clinical colleagues at Nottingham University Hospitals and Liverpool School of Tropical Medicine, have explored the nature of this out of hours work, aiming to understand how technology can be used to support task allocation for clinicians, how rotas can be effectively designed to support out of hours work, and how junior doctors can be trained to prepare them for out of hours shifts.

Several methods are being developed to examine this performance. One is tracking of location. Figures 2 and 3 show illustrative examples of the types of data that can be collected by combining location tracking with observation of activity (captured via expert observers using time-stamped tools on tablet computers). These data, collected from a sample of ~30 hours of observations of junior doctors, show that we can begin to make inferences about the types of requirements of space within clinical contexts. For example, it can be seen in Figure 2 that talking face-to-face took place in one of three contexts – the nurses’ station, office or stores. The fact that these conversations took place in the stores at first seems counter-intuitive, but is likely to suggest the need for private spaces, where sleeping patients will not be disturbed, for conversations to take place. It may also suggest the level of busyness experienced by a clinician – they are not able to
separate time for searching for items in the stores and talking to colleagues; instead they need to multi-task, by communicating whilst searching for tasks or conducting some point-of-care testing activities.

Figure 2. Data comparing physical location and observed activity (size of dot indicates proportion of time spent on task).

Figure 3. Further example of data comparing physical location and observed activity.
Figure 4 shows data collected from a different source – the system used by hospitals to allocate out-of-hours tasks to clinicians via smartphone devices (see Pinchin et al., 2014). This shows an overview of number of tasks by time of day, and quickly demonstrates the potential value of repurposing these embedded technologies for the purposes of E/HF insight. More detailed analysis of task type, time of specific job allocation, and time when job is noted as complete, has the potential to provide even more detailed analysis of clinical work and develop case studies to support clinical training in task prioritisation. When combined with the location and activity data as shown in figures 2 and 3, these data also allow inference and prediction of demand of different types of tasks on clinical staff; this can then provide the basis of supporting ‘demand-driven staffing’ that ensures that the ideal number and type of staff are allocated to different types of day (of course, it is acknowledged that the complex systems-nature of the clinical organisational and business context makes this argument easier in principle than in practice!).

3. Methodological considerations

The case studies have provided illustrations of the different potential value of application of ubiquitous technologies for capture of real-world performance; there are some methodological considerations in the implementation and analysis of such technologies that should be considered before they can be adopted as part of the E/HF practitioner’s toolkit.

3.1 Data granularity

When applying technologies to capture performance, decisions need to be made regarding the appropriate level of data granularity. Whereas in the past it was often the case that an E/HF practitioner was frustrated at the lack of sensitivity or capacity of a data collection tool (e.g. for measuring environment in the workplace) the ubiquitous nature, increased storage and sensitivity of tools, and relative ease of use of technologies now often presents a temptation to capture data at a much higher level than is actually needed. Therefore it is both practically and ethically important to decide on the required level of granularity in advance of data collection. The practical value of this is that it makes data collection (e.g. sampling of data of interest around a specific interaction or event) more efficient and less costly; the ethical value is actually associated with the potential liability of collecting data but for reasons of time or financial capacity not actually analysing that data to the level of detail that it was collected. It may often be the case when using bespoke technologies that the
researcher is not actually in a position to alter the level of granularity of data collected; if this is the case, then the participant should be informed of the type and detail of analysis to be performed in advance of consenting to the study.

3.2 Intrusiveness of measures

Although methods are often described as ‘ubiquitous’ or ‘pervasive’, and are increasingly either embedded within our environment, or so frequently encountered that they might be considered ‘normal’ (for example, using an accelerometer from within an individual’s personal mobile phone to collect data about their movement around a workplace), it is still important for E/HF practitioners to be aware of the way in which performance changes as a result of applying performance measurement technologies. For example, the eye tracking glasses deployed within case study 1 were light and comfortable, and participants did not report any discomfort or change in behaviour as a result of wearing them. In part of the research programme reported in case study 2, clinicians willingly wore additional tracking devices that captured their movements around the hospital during their shifts, in the form of small mobile devices encased in bespoke cases.

However, despite the clear ‘acceptance’ of both technology and participation in the study, demonstrated by participants in both case studies, there is still a need to understand whether, and how, behaviour is altered as a result of the awareness that performance is being monitored. Both cases considered are in safety critical environments where it is routinely expected that colleagues and systems will record performance; it may well be the case that in other work contexts the technology is not as easily accepted, and that the behaviour of participants may change to a greater extent as a result of the data collection being undertaken.

3.3 Construct and content validity

Validity refers to the notion of whether a measure measures what it purports to measure, and content and construct validity are particularly concerned with whether a method is sufficient, relevant and complete (see Wilson & Sharples, 2015). It is important when presented with an embedded technology measure of performance, such as with the smartphone data system that is used to allocate jobs to clinicians in hospitals, to understand the validity of these data.

Our work demonstrates that the mapping between the data collected and the inferences made about activity may not be straightforward; for example, it is easy to note from the system the time at which the job was allocated, and the time at which it was marked complete. What is not so straightforward is to understand when, from an experienced workload perspective, the task began to exert demand on the individual. For example, a doctor may accept a job in order to signal to the nurse that they are happy to take on responsibility for completing that task. There may be a time delay before the onset of starting the task, perhaps due to the need to complete another task, to attend to personal needs, or simply to travel to the location where the task can be started (although in the case of a complex task, doctors have anecdotally reported that they begin thinking about how to complete the task before they physically arrive at the relevant ward). The same complexity applies to interpretation of job completion data; our observations and discussions with clinicians reveal that often doctors choose not to note a task as completed, and instead retain it on their list of active jobs. This might, for example, help to remind them to check in with the patient later on in their shift, to mention a point to a colleague during handover, or to chase up some results. This is indicative of an ‘attenuation’ of a task, when it is no longer the primary demand on their time, but neither is it considered complete. All of this is completely acceptable within the context of the job of a clinician, but needs to be appreciated when interpreted the data as indicators of performance.

3.4 Face validity

Face validity is clearly related to, but distinct from, content and construct validity. Face validity refers to whether a measure looks like it measures what it purports to measure, and can be extremely important when using E/HF data to influence change or investment in workplace or system design. This is also relevant to the application of technologies in a real world context.

Ubiquitous technologies have great potential to present a persuasive argument for workplace change. Whilst those of us with E/HF methodological expertise will, rightly, have faith in appropriately applied well established approaches that use methods such as expert analysis, ethnographic observation or self report, it
can, rightly or wrongly, be harder to convince someone to make a workplace change on the basis of data that they perceive as ‘subjective’ (see Wilson & Sharples (2015) for a longer debate on the notion of subjectivity in E/HF methods). Methods such as eye tracking have high face validity – a non expert can quickly see the link between the workplace with which they are familiar, and the reported data. It is interesting that this high face validity is not always as straightforward as it might at first appear (for example, there is much discussion about the relationship between eye movements and notions of visual attention (Jacob & Karn, 2003)) but it is important to acknowledge the ‘persuasive power’ of a method with high face validity.

3.5 Complementarity and data triangulation

An aspiration of many research activities is to achieve ‘data triangulation’, where we measure the same phenomenon from different perspectives. It can sometimes be a challenge to introduce truly independent methods into an individual or team, so any approach that increases the potential number of methods which can be applied, without compromising the real world context of data collection, is useful to explore. Both cases illustrate the first steps towards data triangulation: in case 1, eye movement analysis is combined with verbal protocol and video analysis of performance (e.g. to identify when critical incidents, such as approaches to level crossings, occur); in case 2, observed activity data are combined with location and task data.

3.6 Ethics

Finally, but probably most importantly, the ethics of collecting data of this detail in a real world workplace need to be carefully considered. Ethics protocols need to not only consider the routine matters of consent, but the ways in which data are stored, to ensure that privacy of individuals is protected. The level of detail within data has the potential to make individuals retrospectively identifiable (e.g. a doctor may well be able to pick out which was his/her shift from a selection of data patterns). There are also clear potentials for these data which are collected for the purposes of E/HF interventions and analysis to also be repurposed for performance assessment – for example, a train driver’s visual strategies could be analysed, or a clinician could be assessed based on the number of steps walked during a shift. Such uses of data must be anticipated wherever possible, and appropriate safeguards put in place to prevent misuse of data.

4. Conclusions

Ubiquitous technologies have clear potential to enhance the way in which we collect data about workplace behaviour and performance. They are not, however, a panacea, and we must carefully consider the most appropriate ways in which such tools should be implemented within workplace contexts, and how their data should be used. If measures are taken to ensure that the rights and concerns of individuals who are interacting with these technologies are protected, then technology based tools that are embedded within workplaces over extended periods of time have the potential to enhance our understanding of real world behaviour and performance, and make more effective design decisions and interventions in future workplaces.

Acknowledgements

The work reported within this paper was carried out in collaboration with Horizon Digital Economy Research, through the support of RCUK grant EP/G065802/1, the MATCH project through the support of EPSRC grant EP/F063822/1 and Network Rail.

References


