Highlights

Using IoT Devices for Sensor-based Monitoring of Employees' Mental Workload: Investigating Managers' Expectations and Concerns

Sebastian Pütz,Vera Rick,Alexander Mertens,Verena Nitsch

- Most managers expect improvements for employee well-being and workplace design
- Bayesian models show associations of expectations and concerns with system support
- Privacy concerns are the main barrier to the adoption of mental workload monitoring
- Managers from Germany, the UK, and Spain differ in support for workload monitoring

Using IoT Devices for Sensor-based Monitoring of Employees' Mental Workload: Investigating Managers' Expectations and Concerns

Sebastian Pütz[a,*], Vera Rick[a], Alexander Mertens[a] and Verena Nitsch[a,b]

[a] Institute of Industrial Engineering and Ergonomics, RWTH Aachen University, Eilfschornsteinstr. 18, 52062 Aachen, Germany
[b] Fraunhofer Institute for Communication, Information Processing and Ergonomics FKIE, Campus-Boulevard 55-57, 52074 Aachen, Germany

ARTICLE INFO

Keywords:
Mental workload
Monitoring
Privacy concerns

ABSTRACT

Although the objective assessment of mental workload has been a focus of human factors research, few studies have investigated stakeholders' attitudes towards its implementation in real workplaces. The present study addresses this research gap by surveying N = 702 managers in three European countries (Germany, United Kingdom, Spain) about their expectations and concerns regarding sensor-based monitoring of employee mental workload. The data confirm the relevance of expectations regarding improvements of workplace design and employee well-being, as well as concerns about restrictions of employees' privacy and sovereignty, for the implementation of workload monitoring. Furthermore, Bayesian regression models show that the examined expectations have a substantial positive association with managers' willingness to support workload monitoring in their company. Privacy concerns are identified as a significant barrier to the acceptance of workload monitoring, both in terms of their prevalence among managers and their strong negative relationship with monitoring support.

1. Introduction

Since its popularization in the late 1970s (Moray, 1979), the construct of mental workload has become one of the most widely used constructs in human factors research (Young et al., 2015). It is usually defined as the proportion of an individual's cognitive resources that must be expended to perform a given task under specific environmental and operational conditions (e.g. Cain, 2007; Curry et al., 1979; Gopher and Donchin, 1986). In cases of mental overload or underload, the individual's ability to cope with task demands is impaired, resulting e.g. in slower working speeds including slower reactions to critical events and higher error rates (Sharples and Megaw, 2015). The goal of designing a workplace with the the employee's mental workload in mind is, thus, to prevent performance degradation due to overload or underload potentially resulting in both a loss of productivity and workplace safety as well as negative consequences for the individual's mental well-being such as increasing mental fatigue (Fan and Smith, 2017; Grech et al., 2009) or mental stress (Cinaz et al., 2013; Gaillard, 1993).

While the large body of research on factors influencing mental workload in laboratory settings provides valuable insights for workplace and system design, these approaches typically fall short in capturing all relevant interdependencies in the face of the multifaceted and dynamic nature of real work environments that leads to constant fluctuations of external demands and internal resources. Consequently, an increasing amount of research strives for the application of mental workload assessment in real workplaces enabling monitoring employees' workload. This would facilitate the detection and prevention of inappropriate workload, especially in non-standardized and safety-critical work environments (cf. van Acker et al., 2020a).

Although mental workload monitoring in real work environments is an ambitious goal, there are two main developments that can provide the basis for its realization. The first one is the advancement in psychophysiological measures of mental workload. Beside performance-based measures and subjective rating scales, psychophysiological approaches have long been investigated as a proxy for mental workload (Eggemeier et al., 1991). Their main strength is that they can be applied continuously during task execution, providing an objective assessment of the individual's workload (Charles and Nixon, 2019). While early research focused on investigating the validity of these approaches and their use in laboratory settings, a rising number of studies evaluates their applicability in the field across a variety of industries (e.g. Fallahi et al., 2016; Kennedy-Metz et al., 2021; Scannella et al., 2018; Szewczyk et al., 2020).

The second development is the deployment of internet of things technology (IoT; Xia et al., 2012). IoT devices combine embedded sensors and processing capacities with strong interconnectivity enabling a distributed network of devices. This developing digital infrastructure can be used or extended to facilitate workload monitoring. For example, wrist-worn wearables, which are introduced in construction to improve workplace safety (Ahn et al., 2019; Awolusi et al., 2018; Barata and da Cunha, 2019), can measure common psychophysiological workload indices. Another industry that records a steep increase in the utilization of IoT technology is manufacturing (Qu et al., 2016; Zhang et al., 2016). Although most research efforts are focused on using gathered data on machines and technical processes, a current impetus is to extend this practice in line with the concept of workload monitoring to the human factor in sociotechnical production systems (Mertens et al., 2021).

Considering that major research efforts have been made to develop methods for monitoring the mental workload of

*Corresponding author
s.puetz@iaw.rwth-aachen.de (S. Pütz)
ORCID(s): 0000-0002-2948-3953 (S. Pütz)

workers in real workplaces, there is surprisingly little research on whether employers and employees would actually support the introduction of such measures (cf. van Acker et al., 2020a). Despite the aforementioned advantages, the introduction of respective technologies in the workplace may also have negative consequences that could outweigh the benefits for relevant stakeholders. To address this research gap, this study takes a closer look at potential expectations and concerns towards employee workload monitoring from members of the industry. Investigating stakeholder attitudes is essential as they constitute important predictors of future behavior (Ajzen, 1991; Venkatesh and Davis, 2000). They will therefore determine whether stakeholders will support or oppose the adoption of workload monitoring, shaping the systems' potential for success.

2. Literature review

2.1. Objective mental workload assessment

The following section provides an overview of the current state of research on objective workload assessment methods that do not rely on subjective self-ratings and could enable continuous workload monitoring. Beside demonstrating the variety of approaches, the goal is to illustrate the practical orientation of the corresponding research by focusing on applications for different industries, highlighting implementations in the field.

As introduced, most methods suitable for mental workload monitoring in practice are based on psychophysiological measures. Among them, measures that assess workload based on electrocardiac activity such as heart rate (HR) or heart rate variability (HRV) are the most prevalent (Charles and Nixon, 2019). In laboratory settings, they have been used across a wide range of domains such as air traffic control (e.g. Kutilek et al., 2018; Metzger and Parasuraman, 2001; Radüntz et al., 2020a), aviation (e.g. Grassmann et al., 2017; Mansikka et al., 2016; Zheng et al., 2019), manufacturing (Rajavenkatanarayanan et al., 2020), military (Matthews et al., 2015), power plant operation (e.g. Gan et al., 2020; Gao et al., 2013; Reinerman et al., 2020), police (Tiwari et al., 2020), process control (Sauer et al., 2013), remote operation (e.g. Durantin et al., 2014; Heard and Adams, 2019; Landi et al., 2018) and ship navigation (Kitamura et al., 2016; Murai et al., 2008; Sugimoto et al., 2016).

In addition to these studies investigating the relationship between electrocardiac activity and mental workload in the laboratory, this approach also accounts for most research in the field with studies conducted in aviation (Noel et al., 2005; Scannella et al., 2018; Wilson, 2002), agricultural operation (Dey and Mann, 2010), city traffic control (Fallahi et al., 2016), office work (Myrtek et al., 1999), surgery (Kennedy-Metz et al., 2021) and train operation (Myrtek et al., 1994). This collection of application domains is even further extended when also considering studies focusing on mental stress (Gaillard, 1993) rather than workload.

In contrast to the plethora of research on electrocardiac measures in applied settings, most other psychophysiological measures are mainly studied in controlled environments. This is the case for the second most reported approach, measuring the electrical brain activity with electroencephalography (EEG) to assess mental workload (Charles and Nixon, 2019). Although EEG based methods have been used in domains such as air traffic control (e.g. Aricò et al., 2019; Bernhardt et al., 2019; Radüntz et al., 2020b), aviation (e.g. Blanco et al., 2018; Hebbar et al., 2021; Lee et al., 2020), construction (Chen et al., 2016, 2017), manufacturing (Argyle et al., 2021), power plant operation (Reinerman-Jones et al., 2016), remote operation (Durantin et al., 2014; Rojas et al., 2020), ship handling (Liu et al., 2020) and software development (Fritz et al., 2014), field studies have mostly been limited to the field of aviation (Dehais et al., 2019; Noel et al., 2005; Wilson, 2002).

A similar research focus can be observed for other common psychophysiological measures such as assessing mental workload based on changes in electrodermal activity (EDA), respiration rate or ocular measures which include variations in pupil size as well as eye movements and blink behavior (Charles and Nixon, 2019). Even though all of them have long been investigated for a variety of application areas (EDA: e.g. Kosch et al. 2019; Setz et al. 2009; Fritz et al. 2014 , respiration: e.g. Argyle et al. 2021; Ding et al. 2020; Gan et al. 2020, ocular measures: e.g. Seeber and Kerzel 2012; Truschzinski et al. 2018; van Acker et al. 2020c), the body of research in the field is still limited.

Beside these long established methods, there are also some comparatively new approaches. One that has gained interest in recent years is the workload assessment based on speech characteristics (van Puyvelde et al., 2018). Previous areas of application include air traffic control (Cosic et al., 2019; Luig and Sontacchi, 2010), aviation (Heard et al., 2019; Huttunen et al., 2011), military (Vukovic et al., 2019) as well as urban search and rescue (Charfuelan and Kruijff, 2013). As a final note, van Acker et al. (2020b) recently proposed a video-based approach. Based on video recordings, they investigated whether the use of a behavioral coding scheme to classify certain movements in an assembly tasks, such as freezing or repetitions, can enable the identification of mental overload. While moving away from psychophysiology, this approach would also meet the demands of an objective sensor-based assessment that could provide continuous workload monitoring when implemented with automated image analysis for behavioral coding.

2.2. Support in the workplace

As noted above, research that specifically examines the support of workload monitoring by employers and employees is comparatively sparse. Although general technology acceptance literature (e.g. Venkatesh et al., 2016; Venkatesh and Davis, 2000) can provide initial guidance, it lacks the consideration of the concrete opportunities and risks of this specific technological development.

One area of research that can be drawn upon is the literature on employee monitoring and employee privacy (cf. Bhave et al., 2020). If an employer oversteps the personal

Table 1
Expectations and concerns about mental workload monitoring, adapted from Mettler and Wulf (2019).

Expectations	Description
Well-Being	Increased awareness of the negative effects of certain work practices on the health and well-being of employees, allowing them to be better addressed
Working Conditions	Better identification and correction of a poor adaptation of working conditions and tasks to employees
Occupational Safety	Earlier detection and prevention of occupational safety risks caused by employee overload
Behavioral Change	Possibility of providing incentives for employees to change unhealthy work practices through open communication of individual or group workload indicators
Concerns	**Description**
Tech. Dependency	Reduction of employees' self-reliance and responsibility at work due to an increasing dependence on technology
Data Sovereignty	Interference with employees' sovereignty over sensitive personal data by measuring their mental workload
Performance Reviews	Compromising the procedural fairness and situational adequacy of job performance reviews
Personal Freedom	Limiting employees' opportunities to pursue self-contained and innovative decisions at work

boundaries of an employee by collecting personal information, the employee will perceive this action as an invasion of privacy negatively impacting their morale (Bhave et al., 2020). The risk of privacy invasions is particularly high with any type of employee monitoring which can become increasingly prevalent with the introduction of sensor data-based assistance systems such as the proposed workload monitoring (Backhaus, 2019). In this context, Backhaus indicated in his meta-analysis that employee monitoring leads to, among other things, an increase in subjective stress and negative affect as well as a decrease in job satisfaction and perceived control.

Another relevant line of research investigates factors influencing the acceptance of IoT devices and, in particular, of wearables in the workplace. Wearables as a reference have the advantage that they are already becoming established in workplaces and that they collect health data with a similar criticality to individual workload. Further, they could indeed be used to obtain workload indices such as HRV or EDA. Surveying 1273 employed adults in the USA, Jacobs et al. (2019) showed a focus on improving workplace safety, a positive safety climate, evidence for validity and involvement of employees in the implementation process as most important for wearable acceptance. Using a different approach, Schall et al. (2018) asked 952 occupational safety and health professionals about their single biggest concern regarding the use of wearables in the workplace. Employee privacy concerns were stated most frequently, followed by concerns about employee compliance, sensor durability, cost-benefit ratio, and distraction from work. Privacy concerns were also mentioned in two studies in the construction industry by Choi et al. (2017) and Häikiö et al. (2020).

Similar hindrances to those mentioned have been discussed by e.g. Ahn et al. (2019), Awolusi et al. (2018), Khakurel et al. (2018), as well as Nappi and de Campos Ribeiro (2020). Based on their literature review, Mettler and Wulf (2019) combined existing literature on wearable acceptance in the workplace into four major affordances and four major constraints. Although initially developed for wearables, these eight aspects, with minor revisions, provide a viable basis for investigating expectations and concerns about mental workload monitoring (s. Table 1).

Only few studies have specifically examined attitudes towards mental state monitoring. In a small study by Brouwer et al. (2018) on mental state monitoring in office environments, concerns mentioned by participants included privacy, validity, and benefits compared to introspection. Furthermore, van Acker et al. (2020a) investigated how acceptance of mental workload monitoring is influenced by framing characteristics in corporate communication.

Overall, the existing literature provides a sound basis for potential expectations and concerns about monitoring technologies, with particular emphasis on the critical role of employee privacy. Whether these findings can be applied to mental workload monitoring remains to be explored.

2.3. Research objective

There is a striking imbalance between the research effort invested into mental workload monitoring and into the requirements for its application in real workplaces. This study aims to contribute to the latter research area by examining the prevalence and role of expectations and concerns about mental workload monitoring among managers. Managers are a fitting target group as they will be (at least partially) responsible for deciding whether respective technologies will be implemented in their company, and they have not been a research focus in this area so far (cf. Schall et al., 2018). By examining specific expectations and concerns about workload monitoring, it is possible to assess whether those identified for IoT devices in general are transferable, whether aspects need to be added, and how these affect managers' willingness to support the use of workload monitoring in their companies. The study also considers the variety of approaches for objective sensor-based workload monitoring, whereby the focus is not to identify a best solution but to

Table 2
Scenarios of sensor-based mental workload monitoring systems included in this study.

System	Description
Environ. sensors	Environmental sensors that measure environmental conditions such as loudness, humidity, or illumination to determine their effect on the employee's mental workload
Wearable	Wearables like fitness trackers that use physiological parameters, such as blood pressure or heart rate, to determine the employee's mental workload
Microphone	Microphone-based systems that use voice characteristics, such as pitch or rate of speech, to determine the employee's mental workload
Camera	Camera-based systems that use body posture or the occurrence of certain behaviors to determine the employee's mental workload
Eye-tracker	Eye-tracker based systems that use eye parameters, such as pupil dilation or eye movement, to determine the employee's mental workload

examine the relevance of approach-specific concerns. The main research questions are therefore:

RQ1: Which expectations and concerns do managers have regarding sensor-based mental workload monitoring?

RQ2: How do these expectations and concerns relate to managers' willingness to support the use of sensor-based mental workload monitoring in their company?

The research questions are examined using a sample of managers from three different European countries. The sample is thus more diverse than in most previous studies in this domain, opening up the possibility of also gaining preliminary insights into potential regional or cultural differences.

3. Methods

3.1. Sample

Participants were recruited from May 3rd to 31st, 2021. The recruitment took place in three European countries: Germany, the United Kingdom (UK), and Spain. The German sample was planned as the main sample of the survey and was targeted to account for two-thirds of the total sample. Participants were contacted via a survey panel provider that accessed a random sample of panel members who met the inclusion criteria for the survey. Inclusion criteria required respondents to be at least 18 years of age and to be employed full or part time as managing directors, middle managers, employees from the strategy department or heads of the training department in companies with at least five employees. Only responses of participants who completed the survey to the end were accepted. Respondents received monetary compensation for their participation.

To ensure sufficient data quality for the following analysis, the data collection was followed by a multi-step screening of the obtained responses. First, participants with implausible completion times were excluded, using the relative speed index suggested by Leiner (2019) with a lenient cut-off of 2.0 as criterion. Second, an attention check item ("*I am currently filling out a questionnaire.*") was included which participants had to pass (Shamon and Berning, 2020). Third, answers were screened for "straightlining" and, finally, answers to open-ended questions were checked for signs of automated or disingenuous responses (standard responses not related to the subject of the questionnaire).

3.2. Questionnaire

After an introductory section which captured participants' demographics, their form of employment, and information about their company, participants were introduced to the concept of sensor-based monitoring of employees' mental workload. For this purpose, participants were given descriptions of five scenarios of workload monitoring systems which are presented in Table 2. The scenarios which are defined by sensor-type were selected to group existing approaches in a way meaningful to the participants and to enable the investigation of the impact of scenario-specific concerns on participants' support of the respective system. For the latter reason, simple environmental sensors were included as a reference scenario. Since environmental conditions, such as loudness, temperature, or illumination, affect humans' capacities to perform cognitive tasks (e.g. Banbury et al., 2001; Banbury and Berry, 2005; Mills et al., 2007; Lan et al., 2010), monitoring them would facilitate the prevention of negative impacts on mental workload. However, they would not allow an assessment of individual workload or collect other personalized data, making them a less concerning reference. The four remaining scenarios were wearables, microphones, cameras and eye-trackers, whereby multiple psychophysiological approaches like HRV, EDA or lightweight EEG systems can be subsumed under the wearables scenario.

In line with the study by Schall et al. (2018), participants could indicate their single biggest concern about the use or the individual systems in their company. Subsequently, participants rated the five scenarios according to three predetermined concerns, namely invasion of employees' privacy, distraction of employees from their work and a lack of employee compliance, on 5-point ordinal scales (*strongly disagree* to *strongly agree*) respectively. The three concerns were selected based on the results of Schall et al. (2018) as the most prevalent concerns among occupational safety and health professionals which could differentiate the scenarios based on the current knowledge of the participants. Participants then assessed sensor-based mental workload monitoring in general based on the four expectations and four con-

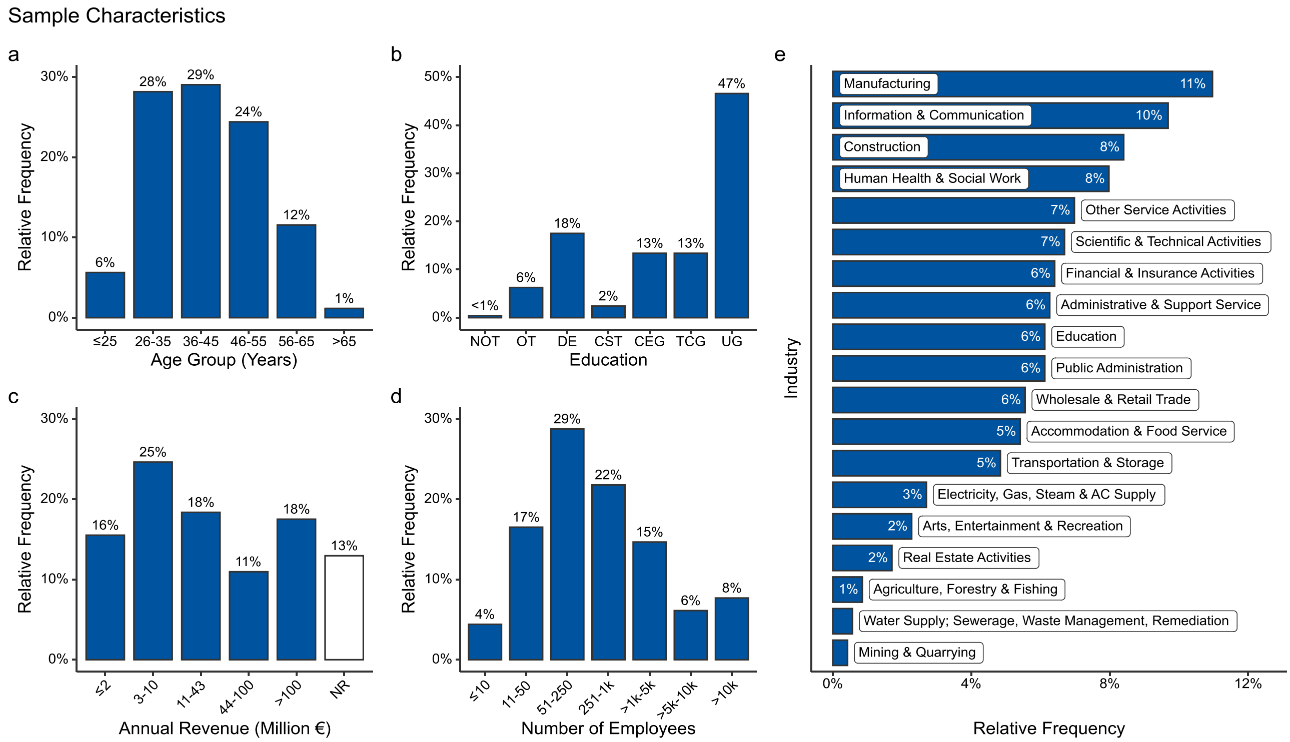

Figure 1: Distributions of sample characteristics: (a) participant's age, (b) participant's education (NOT: No occupational training, OT: Occupational-school training, DE: Dual education, CST: Civil service training, CEG: University of cooperative education graduate, TEG: Technical college graduate, UG: University graduate), (c) annual revenue of the company (NR: Not reported), (d) number of company employees and (e) industry of the company following the NACE (European Commision, 2006).

cerns adapted from the literature review of Mettler and Wulf (2019). Finally, participants were asked for each scenario whether they would support the use of workload monitoring in their company, using 4-point ordinal scales (*no*, *rather no*, *rather yes*, *yes*).

The described questionnaire was part of a larger survey concerning the digitalization within participants' companies. Participants received the questionnaire in the main language of their country.

3.3. Data analysis

To address research question two and investigate the relationship between managers' expectations and concerns and them supporting the application of mental workload monitoring systems, we used Bayesian regression modeling (for introductions see e.g. Kruschke, 2015; McElreath, 2020). In a Bayesian model, existing knowledge, which is specified in form of prior probability distributions for the model parameters, is combined with the obtained insights from observed data to update the knowledge resulting in a joint posterior probability distributions for model parameter values. The posterior can then be used for statistical inference, e.g. by estimating the central tendency and spread of the marginal posterior distribution for individual parameters (for a summary of the advantages of Bayesian inference compared to traditional approaches see e.g. Kruschke and Liddell, 2018; Wagenmakers et al., 2018).

Bayesian linear regression models were fitted with the *R* package *brms* (Bürkner, 2017, 2018) which provides an interface to the probabilistic programming language Stan (Carpenter et al., 2017) applying the No-U-Turn Sampler, an extension to Hamiltonian Monte Carlo (Hoffman and Gelman, 2014). *brms* also provides specific features for handling relevant characteristics of the collected data. First, support for monitoring systems as the outcome variable was assessed on an ordinal scale (cf. Liddell and Kruschke, 2018). Thus, an ordinal regression approach was chosen using a cumulative likelihood and a logit link function to model responses as categorization of a latent continuous variable into *K+1* ordered response categories, separated by *K* thresholds on the latent scale (Bürkner and Vuorre, 2019).

Second, expectations and concerns as model predictors were also ordinal. These were modelled as monotonic effects constraining the effect for each transition between adjacent response categories to have the same direction but allowing the transitions to account for different proportions of the total predictor effect (Bürkner and Charpentier, 2020). Third, multiple responses were collected per participant and per scenario requiring a multilevel model structure. Observations were, therefore, cross-classified modelling differences in average support levels between participants and between systems.

The following priors were set with the goal of ruling out unreasonable parameter values without biasing the estima-

Table 3
Absolute Frequency of single biggest concern types per monitoring system.

	Priv.	Comp.	Dura.	Co./Be.	Dist.	Go.Ma.	Vali.	He.Ri.	Saef.	Usef.	Other
Environ. sensors	49	3	0	6	1	1	5	1	5	4	11
Wearable	84	8	1	5	3	1	1	0	3	7	11
Microphone	112	6	0	3	0	0	3	0	3	4	15
Camera	155	7	0	3	1	3	2	0	8	2	8
Eye-tracker	100	6	0	5	1	0	4	1	5	11	15
Overall	500	30	1	22	6	5	15	2	24	28	60

Note: Priv.: Privacy concerns, Comp.: Employee compliance, Dura.: Sensor durability, Co./Be.: Cost/benefit ratio, Dist.: Distraction from work , Go.Ma.: Good manufacturing practice, Vali.: Validity, He.Ri.: Health risks for employees, Safe.: Safe electronics, Usef.: Usefulness.

Table 4
Descriptive statistics per monitoring system. Means and *SDs* for the system specific concerns (5-point scales) and system support (4-point scale) as well as the relative frequencies of binned support and stated single biggest concerns.

System	Compliance Mean (*SD*)	Distraction Mean (*SD*)	Privacy Mean (*SD*)	Support Mean (*SD*)	Support Frequency	Concerns Frequency
Environ. sensors	3.23 (1.12)	3.06 (1.16)	3.37 (1.20)	3.01 (0.88)	78 %	12 %
Wearable	3.27 (1.14)	3.33 (1.10)	3.51 (1.16)	2.61 (1.00)	58 %	18 %
Microphone	3.41 (1.14)	3.42 (1.16)	3.78 (1.19)	2.38 (1.02)	47 %	21 %
Camera	3.45 (1.10)	3.47 (1.11)	3.96 (1.12)	2.42 (1.01)	47 %	27 %
Eye-tracker	3.37 (1.10)	3.44 (1.11)	3.64 (1.14)	2.38 (0.99)	46 %	21 %
Overall	3.34 (1.12)	3.35 (1.14)	3.65 (1.18)	2.56 (1.01)	55 %	20 %

tion of reasonable values: Normal priors ($\mu = 0$, $\sigma = 2$) for the regression slopes, Student's t priors ($\nu = 3$, $\mu = 0$, $\sigma = 10$) for the threshold parameters of the ordinal outcome variable, non-negative Student's t priors ($\nu = 3$, $\mu = 0$, $\sigma = 2.5$) for the varying intercepts , and Dirichlet priors ($\alpha = 1$) for modelling the structure of the monotonic effects. The adequacy of the implemented priors was also assessed via prior predictive checks (Gabry et al., 2019).

The posteriors of each model were sampled using four Markov chains, each with 2500 iterations before and 5000 iterations after warm-up. Beside diagnostics for model convergence, model adequacy was checked using posterior predictive checks (Gabry et al., 2019) and models were compared via Pareto-smoothed importance sampling leave-one-out cross validation (PSIS-LOO-CV; Vehtari et al., 2017). PSIS-LOO-CV is used to provide an estimate for the expected log pointwise predictive density (*elpd*) of the model, whereby differences in model *elpd* can be interpreted in relation to the standard error (*SE*) of the difference.

For model interpretation, we report the medians of the marginal posterior densities as indicator of effect magnitudes and 95 % Bayesian credible intervals (*CI*) for the precision of the estimate. For those more accustomed to the frequentist p-value, the probability of direction (*pd*) is also reported which describes the proportion of the posterior distribution that has the same sign as the median (Makowski et al., 2019b). In other words, *pd* specifies the probability that an effect is strictly positive/negative. Common thresholds for p-values (0.05, 0.01, 0.001) can be easily mapped to *pd*-values (97.5 %, 99.5 %, 99.95 %; Makowski et al., 2019a).

4. Results

4.1. Sample characteristics

A total of 1183 complete responses were collected that met the inclusion criteria. Of those, 148 cases were removed due to exceeding the specified cut-off for the relative speed index and 159 additional cases for failing the included attention check. 132 responses were excluded for straightlining and, finally, 42 were removed based on screening open answers for automated or disingenuous responses. Thus, the final sample size for further analysis was $N = 702$, which splits into $n_{GER} = 472$, $n_{UK} = 98$, and $n_{SPA} = 132$ for the subsamples. Among the participants are 233 women (33 %), 468 men (67 %) and one participant identifying as divers. 225 participants work as managing directors (32 %), 376 as middle managers (54 %) and 134 as heads of training departments (19 %), whereby 31 participants indicated two of those positions and one participant indicated all three. Further characteristics of the sample are depicted in Figure 1.

4.2. Expectations and Concerns (RQ1)

The following summary of results is based on the full sample. Separate versions of all tables and figures for the subsamples are included in supplementary materials.

To begin with, we analyzed participants' open responses to their single biggest concern regarding workload monitoring systems (s. Table 3). The clustering of the responses

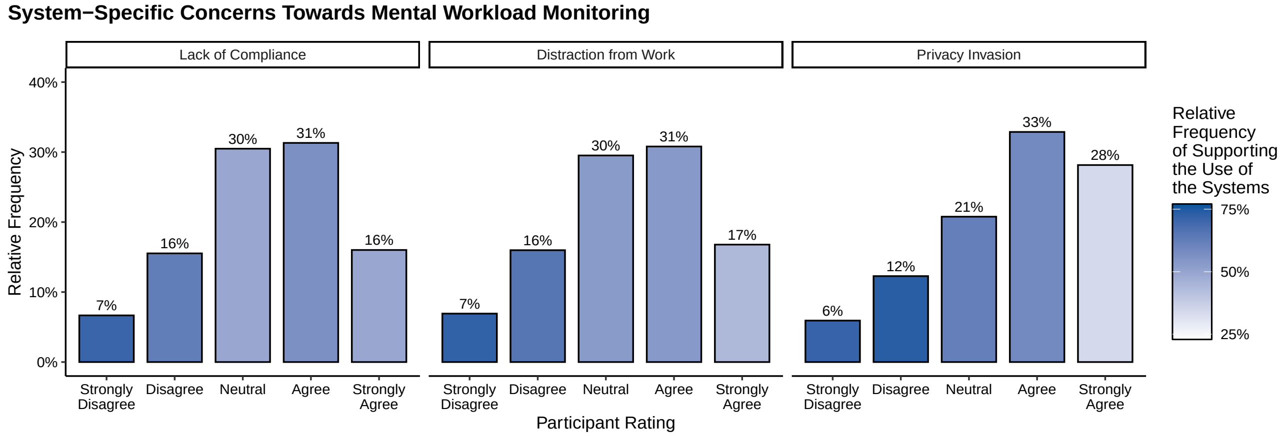

Figure 2: Distributions of participants' ratings of the investigated system specific concerns towards sensor-based mental workload monitoring. The color gradient depicts the relative frequency of binned system support across all five systems for the participants who gave the respective rating for the particular concern, with darker values corresponding to higher frequencies of support.

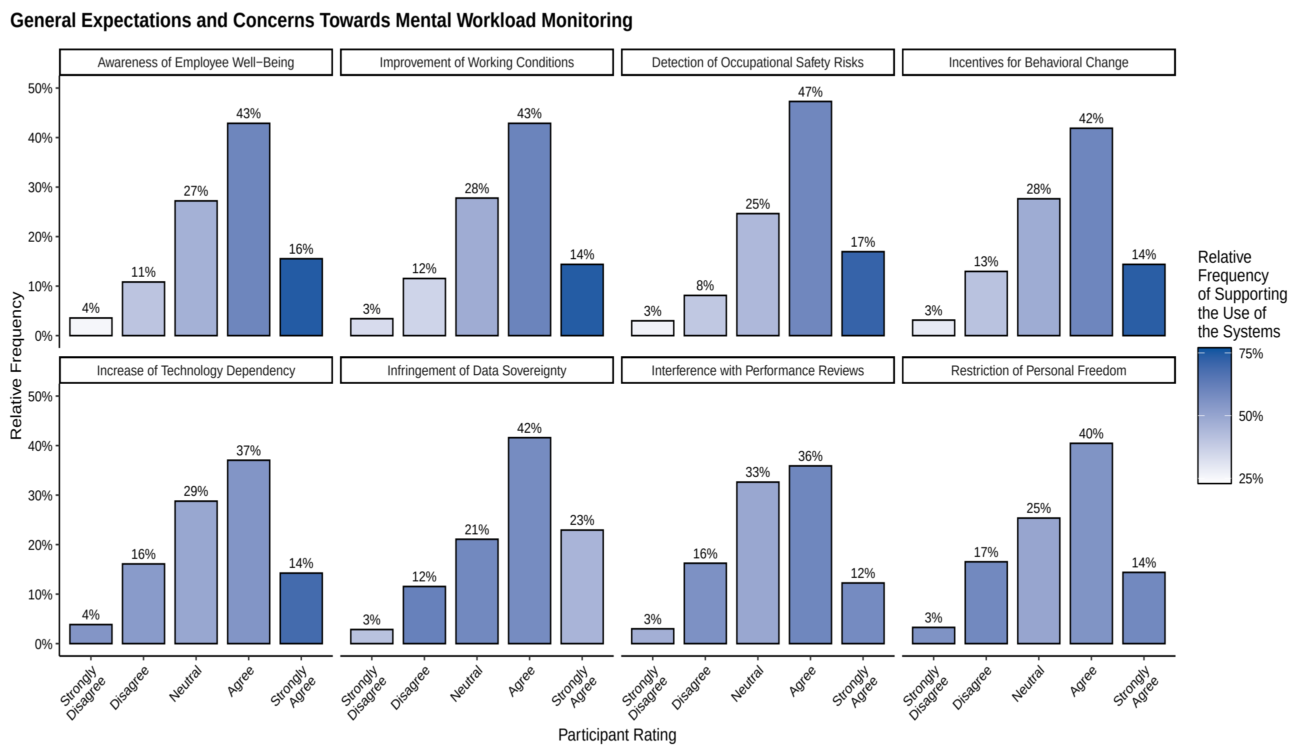

Figure 3: Distributions of participants' ratings of the investigated general expectations (top row) and concerns (bottom row) towards sensor-based mental workload monitoring. The color gradient depicts the relative frequency of binned system support across all five systems for the participants who gave the respective rating for the particular expectation/concern, with darker values corresponding to higher frequencies of support.

was based on the categories identified by Schall et al. (2018) to improve comparability with existing research. The qualitative analysis reveals invasions of employees' privacy as the main concern about workload monitoring systems, as it accounts for 72 % of all stated concerns with 500 respective responses. In contrast, no other concern type makes up more than 30 responses. Focusing on privacy concerns, the camera scenario yields the highest frequency with 38 % more corresponding responses than the microphone scenario, the second scenario in this regard.

Continuing with the quantitative results, Table 4 presents descriptive statistics for the scenario-specific concerns. All three assessed system-specific concerns yield a global mean above the neutral response value of 3, indicating empirical support for their relevance, with privacy invasion again being the most prominent. Furthermore, the binned willing-

Table 5
Posterior median, 95 % *CI* and *pd* (> 97.5 % in bold) for all population-level effects of the final Bayesian regression model with Germany as the reference category for the sample variable.

	Median	95 % CI		*pd*
Sample				
UK	0.916	0.413	1.406	**100.00 %**
Spain	0.523	0.089	0.955	**98.93 %**
General Expectations				
Well-Being	0.260	0.061	0.500	**99.57 %**
Working Conditions	0.374	0.177	0.598	**99.99 %**
Occupational Safety	0.389	0.199	0.637	**100.00 %**
Behavioral Change	0.387	0.150	0.683	**99.98 %**
General Concerns				
Technology Dependency	0.248	0.051	0.414	**98.86 %**
Data Sovereignty	-0.243	-0.441	-0.074	**99.77 %**
Performance Reviews	0.204	0.009	0.444	**97.91 %**
Personal Freedom	-0.162	-0.372	0.076	91.76 %
System-specific Concerns				
Compliance	-0.107	-0.253	0.024	94.31 %
Distraction	-0.259	-0.399	-0.134	**100.00 %**
Privacy	-0.562	-0.696	-0.443	**100.00 %**

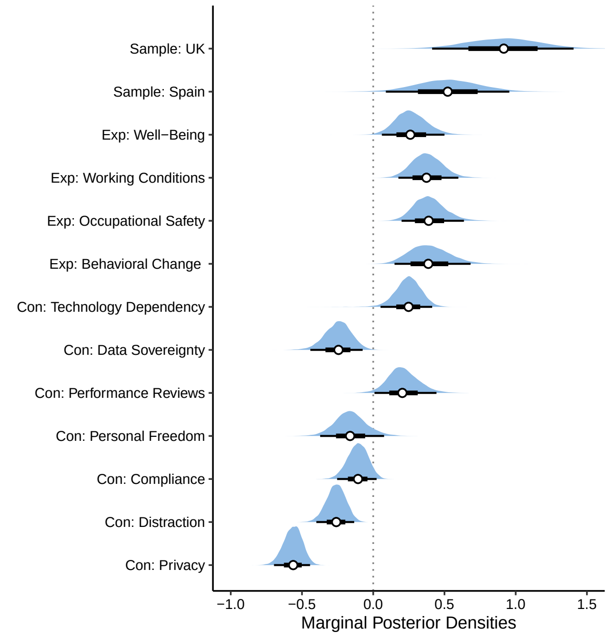

Figure 4: Marginal posterior densities for all population-level effects of the final Bayesian regression model with Germany as the reference category for the sample variable. Point-intervals depict the posterior median as well as 66 % and 95 % *CI*.

ness to support the use of monitoring systems in the own company (*no*, *rather no* vs. *rather yes*, *yes*) shows that the application of respective systems is supported in just above half of the assessments (55 %). Comparing the scenarios, the reference scenario of environmental sensors leads to the least amount of responses addressing participants' biggest concern about using the system and yields the lowest values across all three selected concerns. Whereas the wearable scenario follows second to the environmental sensors, the camera scenario shows the most negative responses with the highest frequency of stated biggest concerns and the highest levels across all three concern types.

Figure 2 and Figure 3 show the ordinal response distributions for system-specific concerns as well as general expectations and concerns towards workload monitoring, respectively. Similar to the system-specific concerns, the relevance of the investigated general expectations is supported by the data since all expectations are shared by the majority of participants. There are no apparent variations in responses comparing the four expectations. In contrast, the investigated general concerns show some differences although in all four cases more participants share the concern than do not. While slightly less than half of the sample (48 %) share the concern that mental workload monitoring would interfere with performance reviews, about two thirds of participants (65 %) perceive workload monitoring as an infringement of employees' data sovereignty.

4.3. Associations with system support (RQ2)

The color gradient in Figure 2 and Figure 3 provides a first look at the associations between the investigated expectations and concerns and participants' support for applying sensor-based mental workload monitoring in their company. The color gradient depicts the relative frequency of the participants who selected the respective response for the particular expectation/concern that would support the use of workload monitoring systems. For system-specific concerns, the color gradient shows a close to monotonic decrease of support among participants with increasing intensity of the respective concern. The most pronounced drop in support exists for those who are highly concerned about privacy invasions. A similar pattern can be seen for all four general expectations which yield strict monotonic increases in support with increasing expectation intensity. Conversely, the relationship between the general concerns and workload monitoring support seem less clear.

To examine the associations, we used Bayesian regression modelling as described in Section 3.3. To begin with, a baseline model for predicting system support was set-up including only the varying intercepts for participants and scenarios as well as a categorical predictor specifying the corresponding subsample with *Germany* as reference. Subsequently, the model was successively extended by first adding monotonic effects for general expectations and concerns and then for system-specific concerns. This order for model expansion was chosen since the decision whether mental workload monitoring is used has to proceed the decision for a specific monitoring system. Comparing the models using PSIS-LOO-CV *elpd* estimates shows that the model fit is improved by both adding general expectations and concerns ($\Delta elpd = 15.2, SE = 6.4$) as well as system-specific concerns ($\Delta elpd = 138.9, SE = 20.8$). Therefore, the estimates of the population-level effects are reported for the full model (s. Table 5 and Figure 4). Estimates are on the scale of the standardized latent variable which was modelled as the basis of the ordinal system support responses.

First of all, the analysis shows significant differences between the subsamples (*pd*s larger 97.5%) which are not explained through the investigated expectations and concerns. Both the UK and the Spain subsample show higher levels of support than the German subsample. The precision of the estimates is comparatively low due to the small sample sizes of these two groups. Moving on to general expectations, all four yield significant positive associations with participants' support for workload monitoring systems (all *pd*s larger 99.5%). In contrast, among the general concerns only the infringement of employees' data sovereignty shows the expected significant negative relationship with system support ($pd = 99.77\,\%$). Whereas the results for restricting employees' personal freedom are not conclusive ($pd = 91.76\,\%$), interference with performance reviews and increase in technology dependency actually yield significant positive associations with system support (both *pd*s larger 97.5%). Finally, the system-specific concerns privacy invasion and distraction show significant negative relationships with system support (both $pds = 100\%$), whereas the negative association of employee compliance concerns is not conclusively supported by the data ($pd = 94.31\,\%$). Consistent with the descriptive analysis, the concern about invasion of employee privacy is the strongest predictor in the model among all the concerns and expectations examined.

5. Discussion

The monitoring of employees' mental workload has been a focus of human factors research over the past decades. Nevertheless, our understanding of the factors influencing employers' and employees' decision to support its implementation is highly limited. Therefore, the conducted study investigated which expectations and concerns are prevalent among managers and how they relate to the managers' willingness to support workload monitoring in their company.

Starting with general expectations towards mental workload monitoring, all four expectations which were adapted from the literature review on wearables by Mettler and Wulf (2019) were shared by the majority of the sample. These results are promising as they suggest that a relevant proportion of managers recognize the benefits that workload monitoring can provide. Even more importantly, all four expectations are significantly positively related to managers' support for workload monitoring. This highlights the four objectives as appropriate for demonstrating the validity of workload monitoring to increase support among corporate decision makers. By stressing the importance of the expected benefits for system acceptance, the results are in line with existing literature both in regard to technology deployment in general (e.g. Venkatesh et al., 2016; Venkatesh and Davis, 2000) and to the introduction of IoT devices in specific (e.g. Choi et al., 2017; Jacobs et al., 2019; Princi and Krämer, 2019; Yassaee and Mettler, 2019).

In contrast to the general expectations, participants' assessments differed to some extent among the four general concerns, with infringement of employee data sovereignty being the most prominent concern. Crucially, the conducted inferential analysis showed that the two concerns of increased technology dependency and interference with performance reviews are in fact positively related with system support. Especially the latter raises concerns about a potential dual use of the workload monitoring data when considering the occupational background of the sample. Although the intended effect of workload monitoring is to improve working conditions for employees to promote their safety, well-being, and performance, the data could also be used to make consequential judgments about employees' performance capacity or stress tolerance (Maltseva, 2020). Effective policies are needed to ensure that promoting employee well-being does not take a back seat to rigorously increasing organizational productivity by using the data to cut or swap personnel (McAleenan et al., 2019).

Regarding system-specific concerns, the negative associations of privacy and distraction concerns with system support were demonstrated. One of the main findings of this study is the major role of privacy concerns, as it was the most prevalent concern in quantitative ratings, by far the most frequent concern in open responses and showed the strongest association with system support. Even though the role of privacy for the acceptance of IoT devices has been previously discussed (e.g. Ahn et al., 2019; Häikiö et al., 2020; Reid et al., 2017; Schall et al., 2018; Yassaee and Mettler, 2019), the estimated extend of this relationship exceeds common reports in the literature highlighting the potential sensitivity of workload data as well as the intrusiveness of sensor-systems such as cameras and microphones.

The problematic nature of individual information privacy is exacerbated in workplaces as there is an inherent power imbalance between employers and employees with employers deciding whether new technological solutions are deployed (cf. Princi and Krämer, 2019; van Acker et al., 2020a). This is also one reason why workload monitoring will face legal difficulties due to European privacy legislation since it is questionable whether workers can give sufficient consent (Collins and Marassi, 2021). Thus, implementing mental workload monitoring will require privacy protection measures such as data anonymization and limiting data accessibility finding a balance between obtaining the intended benefits and preventing opportunities for abuse. It is important that employees do not perceive the workload assessment as a loss of control over personal data (Chen et al., 2013) and that sufficient trust is established in the employer's data processing practices (Chang et al., 2015).

Considering employees' privacy will also play a role in identifying mental workload measures that are appropriate for real-world application. As was expected, the presented monitoring systems differed in their assessments regarding privacy with cameras, microphones and eye-trackers evoking the most concerns. Therefore, respective measures would need to provide substantial performance benefits over physiological methods to warrant their application. Interestingly, even the environmental sensor scenario received medium ratings in regard to privacy. This could be because they were presented in the context of mental workload monitoring or because they are actually more concerning than one might expect. For example, collecting data on employees' potential exposure to dangerous substances over time may be considered sensitive health data (Le Feber et al., 2021).

Finally, the Bayesian regression model showed significant differences in support for workload monitoring systems between participants from different countries that were not explained by the studied expectations and concerns. Although a generalization of this observation should be made cautiously due to the small size of the subsamples, the magnitudes of the estimated effects provide a promising basis for future research efforts. Researchers should examine which regional and cultural differences can explain these differences in support and also expend the investigation to other regions of the world.

5.1. Limitations and Future Research

There are some limitations of the conducted study that must be considered when interpreting the obtained results. First, regarding its scope, the study examined only the perspective of one stakeholder group. However, other stakeholders, particularly employees, must also agree to workload monitoring in order to achieve the expected outcomes. Moreover, the focus was on expectations and concerns about the core concept of workload monitoring leaving out other relevant requirements for monitoring systems such as cost, measurement accuracy or sensor durability.

Second, regarding the applied methods, the variables included in the regression analysis were not assessed using comprehensive scales but single items. Consequently, the estimated relationship can be interpreted more conservatively as the association between participants' manifest responses rather than between latent beliefs. Furthermore, for most of the managers, the study probably represented their first exposure to the issue of workload monitoring. This practice makes it more difficult for participants as they evaluate potential rather than real systems. This might have led to some level of acquiescence as participants may have lacked domain knowledge and specific interest (McClendon, 1991). That said, acquiescence is less likely in a European sample (Smith, 2004) with a high education level (Meisenberg and Williams, 2008; Rammstedt et al., 2010) questioned in an online survey format (Weijters et al., 2008).

Beside the limitations specific to this study, there is a need for more research examining mental workload monitoring in real workplaces. This would not only enable the investigation of the real-world applicability of methods developed in controlled environments and addressing corresponding challenges (Alberdi et al., 2016; Can et al., 2019), but it would also improve the foundation for analyzing stakeholders' attitudes towards such systems. Finally, it could be the basis for needed long-term studies investigating how opinions about workload monitoring change over time of system use (e.g. Gorm and Shklovski, 2016).

6. Conclusion

The presented study is one of the first to examine the expectations and concerns of relevant stakeholders' regarding the implementation of sensor-based mental workload monitoring. The four concrete expectations of increased awareness for employee well-being, improvements of working conditions, detection of occupational safety risks, and potentials for incentivizing behavioral change were demonstrated as strongly associated with managers' willingness to support workload monitoring in their company. In contrast, the concerns examined differed in their relationship with monitoring support, whereby the risk of invading employee privacy emerged as the strongest impediment to managerial support for workload monitoring.

Acknowledgments

We would like to thank our colleagues Mattia Lisa Müller and Clarissa Pascoe for their contributions to the online survey and Anne Henschel for her support in data preparation.

Supplementary Materials

Supplementary materials include separate tables and figures for the descriptive statistic for the three subsamples, a correlation matrix for the investigated expectations and concerns, the visual prior and posterior predictive check of the final Bayesian regression model as well as the wording of all items in the questionnaire: LINK

Declaration of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

The online survey has been conducted within the project WorkingAge, funded by the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement N. 826232. The analysis presented in this study has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2023 Internet of Production – 390621612.

References

Ahn, C. R., Lee, S., Sun, C., Jebelli, H., Yang, K., and Choi, B. (2019). Wearable sensing technology applications in construction safety and health. *Journal of Construction Engineering and Management*, 145(11):03119007.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2):179–211.

Alberdi, A., Aztiria, A., and Basarab, A. (2016). Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review. *Journal of biomedical informatics*, 59:49–75.

Argyle, E. M., Marinescu, A., Wilson, M. L., Lawson, G., and Sharples, S. (2021). Physiological indicators of task demand, fatigue, and cognition in future digital manufacturing environments. *International Journal of Human-Computer Studies*, 145(1):102522.

Aricò, P., Reynal, M., Di Flumeri, G., Borghini, G., Sciaraffa, N., Imbert, J.-P., Hurter, C., Terenzi, M., Ferreira, A., Pozzi, S., Betti, V., Marucci, M., Telea, A. C., and Babiloni, F. (2019). How neurophysiological measures can be used to enhance the evaluation of remote tower solutions. *Frontiers in human neuroscience*, 13:303.

Awolusi, I., Marks, E., and Hallowell, M. (2018). Wearable technology for personalized construction safety monitoring and trending: Review of applicable devices. *Automation in Construction*, 85:96–106.

Backhaus, N. (2019). Kontextsensitive Assistenzsysteme und Überwachung am Arbeitsplatz: Ein meta-analytisches Review zur Auswirkung elektronischer Überwachung auf Beschäftigte. *Zeitschrift für Arbeitswissenschaft*, 73(1):2–22.

Banbury, S. P. and Berry, D. C. (2005). Office noise and employee concentration: Identifying causes of disruption and potential improvements. *Ergonomics*, 48(1):25–37.

Banbury, S. P., Macken, W. J., Tremblay, S., and Jones, D. M. (2001). Auditory distraction and short-term memory: Phenomena and practical implications. *Human factors*, 43(1):12–29.

Barata, J. and da Cunha, P. R. (2019). Safety is the new black: The increasing role of wearables in occupational health and safety in construction. In Abramowicz, W. and Corchuelo, R., editors, *Business Information Systems*, volume 353 of *Lecture Notes in Business Information Processing*, pages 526–537. Springer International Publishing, Cham.

Bernhardt, K. A., Poltavski, D., Petros, T., Ferraro, F. R., Jorgenson, T., Carlson, C., Drechsel, P., and Iseminger, C. (2019). The effects of dynamic workload and experience on commercially available EEG cognitive state metrics in a high-fidelity air traffic control environment. *Applied ergonomics*, 77:83–91.

Bhave, D. P., Teo, L. H., and Dalal, R. S. (2020). Privacy at work: A review and a research agenda for a contested terrain. *Journal of Management*, 46(1):127–164.

Blanco, J. A., Johnson, M. K., Jaquess, K. J., Oh, H., Lo, L.-C., Gentili, R. J., and Hatfield, B. D. (2018). Quantifying cognitive workload in simulated flight using passive, dry EEG measurements. *IEEE Transactions on Cognitive and Developmental Systems*, 10(2):373–383.

Brouwer, A.-M., van de Water, L., Hogervorst, M., Kraaij, W., Schraagen, J. M., and Hogenelst, K. (2018). Monitoring mental state during real life office work. In Ham, J., Spagnolli, A., Blankertz, B., Gamberini, L., and Jacucci, G., editors, *Symbiotic Interaction*, volume 10727 of *Lecture Notes in Computer Science*, pages 18–29. Springer International Publishing, Cham.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using stan. *Journal of Statistical Software*, 80(1):1–28.

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, 10(1):395–411.

Bürkner, P.-C. and Charpentier, E. (2020). Modelling monotonic effects of ordinal predictors in bayesian regression models. *The British journal of mathematical and statistical psychology*, 73(3):420–451.

Bürkner, P.-C. and Vuorre, M. (2019). Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science*, 2(1):77–101.

Cain, B. (2007). *A Review of the Mental Workload Literature*. Defence Research and Development Canada Toronto - Human System Integration Section, Toronto.

Can, Y. S., Arnrich, B., and Ersoy, C. (2019). Stress detection in daily life scenarios using smart phones and wearable sensors: A survey. *Journal of biomedical informatics*, 92:103139.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1):1–32.

Chang, S. E., Liu, A. Y., and Lin, S. (2015). Exploring privacy and trust for employee monitoring. *Industrial Management & Data Systems*, 115(1):88–106.

Charfuelan, M. and Kruijff, G.-J. (2013). Classification of speech under stress and cognitive load in usar operations. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 508–512. IEEE.

Charles, R. L. and Nixon, J. (2019). Measuring mental workload using physiological measures: A systematic review. *Applied ergonomics*, 74:221–232.

Chen, J., Song, X., and Lin, Z. (2016). Revealing the "invisible gorilla" in construction: Estimating construction safety through mental workload assessment. *Automation in Construction*, 63(5):173–183.

Chen, J., Taylor, J. E., and Comu, S. (2017). Assessing task mental workload in construction projects: A novel electroencephalography approach. *Journal of Construction Engineering and Management*, 143(8):04017053.

Chen, X., Ma, J., Jin, J., and Fosh, P. (2013). Information privacy, gender differences, and intrinsic motivation in the workplace. *International Journal of Information Management*, 33(6):917–926.

Choi, B., Hwang, S., and Lee, S. (2017). What drives construction workers' acceptance of wearable technologies in the workplace?: Indoor localization and wearable health devices for occupational safety and health. *Automation in Construction*, 84(9):31–41.

Cinaz, B., Arnrich, B., La Marca, R., and Tröster, G. (2013). Monitoring of mental workload levels during an everyday life office-work scenario. *Personal and Ubiquitous Computing*, 17(2):229–239.

Collins, P. M. and Marassi, S. (2021). Is that lawful? Data privacy and fitness trackers in the workplace. *International Journal of Comparative Labour Law*, 37(1):65–94.

Cosic, K., Popovic, S., Sarlija, M., Mijic, I., Kokot, M., Kesedzic, I., Strangman, G., Ivkovic, V., and Zhang, Q. (2019). New tools and methods in selection of air traffic controllers based on multimodal psychophysiological measurements. *IEEE Access*, 7:174873–174888.

Curry, R. E., Jex, H., Levison, W., and Stassen, H. (1979). Final report of the control engineering group. In Moray, N., editor, *Mental Workload*, pages 235–252. Plenum Press, New York, NY.

Dehais, F., Duprès, A., Blum, S., Drougard, N., Scannella, S., Roy, R. N., and Lotte, F. (2019). Monitoring pilot's mental workload using ERPs and spectral power with a six-dry-electrode EEG system in real flight conditions. *Sensors*, 19(6):1324.

Dey, A. K. and Mann, D. D. (2010). A complete task analysis to measure the workload associated with operating an agricultural sprayer equipped with a navigation device. *Applied ergonomics*, 41:146–149.

Ding, Y., Cao, Y., Duffy, V. G., Wang, Y., and Zhang, X. (2020). Measurement and identification of mental workload during simulated computer tasks with multimodal methods and machine learning. *Ergonomics*, 63(7):896–908.

Durantin, G., Gagnon, J.-F., Tremblay, S., and Dehais, F. (2014). Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behavioural brain research*, 259:16–23.

Eggemeier, F. T., Wilson, G. F., Kramer, A. F., and Damos, D. L. (1991). Workload assessment in multi-task environments. In Damos, D. L., editor, *Multiple Task Performance*, pages 207–216. Taylor and Francis, London.

European Commision (2006). NACE Rev. 2: Statistical classification of economic activites in the european community.

Fallahi, M., Motamedzade, M., Heidarimoghadam, R., Soltanian, A. R., and Miyake, S. (2016). Effects of mental workload on physiological and subjective responses during traffic density monitoring: A field study. *Applied ergonomics*, 52:95–103.

Fan, J. and Smith, A. P. (2017). The impact of workload and fatigue on performance. In Longo, L. and Leva, M. C., editors, *Human Mental Workload: Models and Applications*, volume 726 of *Communications in Computer and Information Science*, pages 90–105. Springer International Publishing, Cham.

Fritz, T., Begel, A., Müller, S. C., Yigit-Elliott, S., and Züger, M. (2014). Using psycho-physiological measures to assess task difficulty in software development. In Jalote, P., Briand, L., and van der Hoek, A., editors, *Proceedings of the 36th International Conference on Software Engineering*, pages 402–413, New York, NY. ACM.

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., and Gelman, A. (2019). Visualization in Bayesian workflow. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(2):389–402.

Gaillard, A. W. (1993). Comparing the concepts of mental load and stress. *Ergonomics*, 36(9):991–1005.

Gan, Y., Dong, X., Zhang, Y., Zhang, X., Jia, M., Liu, Z., and Li, Z. (2020). Workload measurement using physiological and activity measures for validation test: A case study for the main control room of a nuclear power plant. *International Journal of Industrial Ergonomics*, 78:102974.

Gao, Q., Wang, Y., Song, F., Li, Z., and Dong, X. (2013). Mental workload measurement for emergency operating procedures in digital nuclear power plants. *Ergonomics*, 56(7):1070–1085.

Gopher, D. and Donchin, E. (1986). Workload: An examination of the concept. In Boff, K. R. and Thomas, J. P., editors, *Handbook of Perception and Human Performance*, pages 1–49. John Wiley & Sons, Hoboken, NJ.

Gorm, N. and Shklovski, I. (2016). Sharing steps in the workplace. In Kaye, J., Druin, A., Lampe, C., Morris, D., and Hourcade, J. P., editors, *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 4315–4319, New York, NY. ACM.

Grassmann, M., Vlemincx, E., von Leupoldt, A., and van den Bergh, O. (2017). Individual differences in cardiorespiratory measures of mental workload: An investigation of negative affectivity and cognitive avoidant coping in pilot candidates. *Applied ergonomics*, 59:274–282.

Grech, M. R., Neal, A., Yeo, G. B., Humphreys, M., and Smith, S. (2009). An examination of the relationship between workload and fatigue within and across consecutive days of work: Is the relationship static or dynamic? *Journal of Occupational Health Psychology*, 14(3):231–242.

Häikiö, J., Kallio, J., Mäkelä, S.-M., and Keränen, J. (2020). IoT-based safety monitoring from the perspective of construction site workers. *International Journal of Occupational and Environmental Safety*, 4(1):1–14.

Heard, J. and Adams, J. A. (2019). Multi-dimensional human workload assessment for supervisory human–machine teams. *Journal of Cognitive Engineering and Decision Making*, 13(3):146–170.

Heard, J., Fortune, J., and Adams, J. A. (2019). Speech workload estimation for human-machine interaction. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 63(1):277–281.

Hebbar, P. A., Bhattacharya, K., Prabhakar, G., Pashilkar, A. A., and Biswas, P. (2021). Correlation between physiological and performance-based metrics to estimate pilots' cognitive workload. *Frontiers in psychology*, 12:555446.

Hoffman, M. D. and Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1):1593–1623.

Huttunen, K., Keränen, H., Väyrynen, E., Pääkkönen, R., and Leino, T. (2011). Effect of cognitive load on speech prosody in aviation: Evidence from military simulator flights. *Applied ergonomics*, 42:348–357.

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

Kennedy-Metz, L. R., Dias, R. D., Stevens, R. H., Yule, S. J., and Zenati, M. A. (2021). Analysis of mirrored psychophysiological change of cardiac surgery team members during open surgery. *Journal of surgical education*, 78(2):622–629.

Khakurel, J., Melkas, H., and Porras, J. (2018). Tapping into the wearable device revolution in the work environment: A systematic review. *Information Technology & People*, 31(3):791–818.

Kitamura, K., Murai, K., Furusho, M., Wang, Y., Wang, J., and Kunieda, Y. (2016). Evaluation of mixed culture bridge teammates' mental workload using heart rate variability: Simulator-based ship handling. In *2016 IEEE International Conference on Systems, Man, and Cybernetics*, pages 875–879, Piscataway, NJ. IEEE.

Kosch, T., Karolus, J., Ha, H., and Schmidt, A. (2019). Your skin resists: Exploring electrodermal activity as workload indicator during manual assembly. In Panach, I., Vanderdonckt, J., and Pastor, Ó., editors, *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, pages 1–5, New York, NY. ACM.

Kruschke, J. K. (2015). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press, New York, NY, 2nd edition.

Kruschke, J. K. and Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, 25(1):178–206.

Kutilek, P., Hejda, J., Smrcka, P., Krivanek, V., Doskocil, R., Farlik, J., Casar, J., and Klikova, M. (2018). Heart rate variability during long-term air traffic control in cadets. In Herencsar, N., editor, *41st International Conference on Telecommunications and Signal Processing (TSP)*, pages 1–5, Piscataway, NJ. IEEE.

Lan, L., Lian, Z., and Pan, L. (2010). The effects of air temperature on office workers' well-being, workload and productivity-evaluated with subjective ratings. *Applied ergonomics*, 42:29–36.

Landi, C. T., Villani, V., Ferraguti, F., Sabattini, L., Secchi, C., and Fantuzzi, C. (2018). Relieving operators' workload: Towards affective robotics in industrial scenarios. *Mechatronics*, 54(3):144–154.

Le Feber, M., Jadoenathmisier, T., Goede, H., Kuijpers, E., and Pronk, A. (2021). Ethics and privacy considerations before deploying sensor technologies for exposure assessment in the workplace: Results of a structured discussion amongst dutch stakeholders. *Annals of work exposures and health*, 65(1):3–10.

Lee, W., Migliaccio, G. C., Lin, K.-Y., and Seto, E. Y. (2020). Workforce development: Understanding task-level job demands-resources, burnout, and performance in unskilled construction workers. *Safety Science*, 123(1):104577.

Leiner, D. J. (2019). Too fast, too straight, too weird: Non-reactive indicators for meaningless data in internet surveys. *Survey Research Methods*, 13(3):229–248.

Liddell, T. M. and Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, 79:328–348.

Liu, Y., Lan, Z., Cui, J., Krishnan, G., Sourina, O., Konovessis, D., Ang, H. E., and Mueller-Wittig, W. (2020). Psychophysiological evaluation of seafarers to improve training in maritime virtual simulator. *Advanced Engineering Informatics*, 44:101048.

Luig, J. and Sontacchi, A. (2010). Workload monitoring through speech analysis: Towards a system for air traffic control. In *27th Congress of The International Council of Aeronautical Sciences*. International Council of Aeronautical Sciences.

Makowski, D., Ben-Shachar, M., and Lüdecke, D. (2019a). bayestestr: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, 4(40):1541.

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., and Lüdecke, D. (2019b). Indices of effect existence and significance in the Bayesian framework. *Frontiers in psychology*, 10:2767.

Maltseva, K. (2020). Wearables in the workplace: The brave new world of employee engagement. *Business Horizons*, 63(4):493–505.

Mansikka, H., Virtanen, K., Harris, D., and Simola, P. (2016). Fighter pilots' heart rate, heart rate variation and performance during an instrument flight rules proficiency test. *Applied ergonomics*, 56:213–219.

Matthews, G., Reinerman-Jones, L. E., Barber, D. J., and Abich, J. (2015). The psychometrics of mental workload: Multiple measures are sensitive but divergent. *Human factors*, 57(1):125–143.

McAleenan, P., McAleenan, C., Ayers, G., Behm, M., and Beachem, Z. (2019). The ethics deficit in occupational safety and health monitoring technologies. *Proceedings of the Institution of Civil Engineers - Management, Procurement and Law*, 172(3):93–100.

McClendon, M. J. (1991). Acquiescence and recency response-order effects in interview surveys. *Sociological Methods & Research*, 20(1):60–131.

McElreath, R. (2020). *Statistical Rethinking*. CRC Press, Boca Raton, FL, 2nd edition.

Meisenberg, G. and Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences*, 44(7):1539–1550.

Mertens, A., Pütz, S., Brauner, P., Brillowski, F., Buczak, N., Dammers, H., van Dyck, M., Kong, I., Königs, P., Kordtomeikel, F., Rodemann, N., Schaar, A. K., Steuer-Dankert, L., Wlecke, S., Gries, T. Leicht-Scholten ,C., Nagel, S. K., Piller, F. T., Schuh, G., Ziefle, M., and Nitsch, V. (2021). Human digital shadow: Data-based modeling of users and usage in the internet of production. In *2021 14th International Conference on Human System Interaction (HSI)*, pages 1–8. IEEE.

Mettler, T. and Wulf, J. (2019). Physiolytics at the workplace: Affordances and constraints of wearables use from an employee's perspective. *Information Systems Journal*, 29(1):245–273.

Metzger, U. and Parasuraman, R. (2001). The role of the air traffic controller in future air traffic management: An empirical study of active control versus passive monitoring. *Human factors*, 43(4):519–528.

Mills, P. R., Tomkins, S. C., and Schlangen, L. J. M. (2007). The effect of high correlated colour temperature office lighting on employee wellbeing and work performance. *Journal of circadian rhythms*, 5(2):1–9.

Moray, N., editor (1979). *Mental Workload: Its Theory and Measurement*. Plenum Press, New York, NY.

Murai, K., Hayashi, Y., Okazaki, T., Stone, L. C., and Mitomo, N. (2008). Evaluation of ship navigator's mental workload using nasal temperature and heart rate variability. In *2008 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1528–1533. IEEE.

Myrtek, M., Deutschmann-Janicke, E., Strohmaier, H., Zimmermann, W., Lawerenz, S., Brügner, G., and Müller, W. (1994). Physical, mental, emotional, and subjective workload components in train drivers. *Ergonomics*, 37(7):1195–1203.

Myrtek, M., Fichtler, A., Strittmatter, M., and Brügner, G. (1999). Stress and strain of blue and white collar workers during work and leisure time: Results of psychophysiological and behavioral monitoring. *Applied ergonomics*, 30:341–355.

Nappi, I. and de Campos Ribeiro, G. (2020). Internet of things technology applications in the workplace environment: A critical review. *Journal of Corporate Real Estate*, 22(1):71–90.

Noel, J. B., Bauer, K. W., and Lanning, J. W. (2005). Improving pilot mental workload classification through feature exploitation and combination: A feasibility study. *Computers & Operations Research*, 32(10):2713–2730.

Princi, E. and Krämer, N. C. (2019). Acceptance of smart electronic monitoring at work as a result of a privacy calculus decision. *Informatics*, 6(3):40.

Qu, T., Lei, S. P., Wang, Z. Z., Nie, D. X., Chen, X., and Huang, G. Q. (2016). IoT-based real-time production logistics synchronization system under smart cloud manufacturing. *The International Journal of Advanced Manufacturing Technology*, 84:147–164.

Radüntz, T., Fürstenau, N., Mühlhausen, T., and Meffert, B. (2020a). Indexing mental workload during simulated air traffic control tasks by means of dual frequency head maps. *Frontiers in physiology*, 11:300.

Radüntz, T., Mühlhausen, T., Freyer, M., Fürstenau, N., and Meffert, B. (2020b). Cardiovascular biomarkers' inherent timescales in mental workload assessment during simulated air traffic control tasks. *Applied psychophysiology and biofeedback*.

Rajavenkatanarayanan, A., Nambiappan, H. R., Kyrarini, M., and Makedon, F. (2020). Towards a real-time cognitive load assessment system for industrial human-robot cooperation. In *2020 IEEE 29th International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 698–705. IEEE.

Rammstedt, B., Goldberg, L. R., and Borg, I. (2010). The measurement equivalence of big five factor markers for persons with different levels of education. *Journal of research in personality*, 44(4):53–61.

Reid, C. R., Schall, M. C., Amick, R. Z., Schiffman, J. M., Lu, M.-L., Smets, M., Moses, H. R., and Porto, R. (2017). Wearable technologies: How will we overcome barriers to enhance worker performance, health, and safety? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 61(1):1026–1030.

Reinerman, L., Mercado, J., Szalma, J. L., and Hancock, P. A. (2020). Understanding individualistic response patterns when assessing expert operators on nuclear power plant control tasks. *Ergonomics*, 63(4):440–460.

Reinerman-Jones, L., Matthews, G., and Mercado, J. E. (2016). Detection tasks in nuclear power plant operation: Vigilance decrement and physiological workload monitoring. *Safety Science*, 88:97–107.

Rojas, R. F., Debie, E., Fidock, J., Barlow, M., Kasmarik, K., Anavatti, S., Garratt, M., and Abbass, H. (2020). Electroencephalographic workload indicators during teleoperation of an unmanned aerial vehicle shepherding a swarm of unmanned ground vehicles in contested environments. *Frontiers in neuroscience*, 14:40.

Sauer, J., Nickel, P., and Wastell, D. (2013). Designing automation for complex work environments under different levels of stress. *Applied ergonomics*, 44:119–127.

Scannella, S., Peysakhovich, V., Ehrig, F., Lepron, E., and Dehais, F. (2018). Assessment of ocular and physiological metrics to discriminate flight phases in real light aircraft. *Human factors*, 60(7):922–935.

Schall, M. C., Sesek, R. F., and Cavuoto, L. A. (2018). Barriers to the adoption of wearable sensors in the workplace: A survey of occupational safety and health professionals. *Human factors*, 60(3):351–362.

Seeber, K. G. and Kerzel, D. (2012). Cognitive load in simultaneous interpreting: Model meets data. *International Journal of Bilingualism*, 16(2):228–242.

Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., and Ehlert, U. (2009). Discriminating stress from cognitive load using a wearable eda device. *IEEE transactions on information technology in biomedicine: a publication of the IEEE Engineering in Medicine and Biology Society*, 14(2):410–417.

Shamon, H. and Berning, C. C. (2020). Attention check items and instructions in online surveys: Boon or bane for data quality? *Survey Research Methods*, 14(1):55–77.

Sharples, S. and Megaw, T. (2015). The definition and measurement of human workload. In Wilson, R. J. and Sharples, S., editors, *Evaluation of Human Work*, pages 515–548. CRC Press, Boca Raton, FL.

Smith, P. B. (2004). Acquiescent response bias as an aspect of cultural communication style. *Journal of Cross-Cultural Psychology*, 35(1):50–61.

Sugimoto, I., Kitamura, K., Murai, K., Wang, Y., and Wang, J. (2016). Study on relation between operator and trainee's mental workload for ship maneuvering simulator exercise using heart rate variability. In *2016 IEEE International Conference on Systems, Man, and Cybernetics*, pages 768–772, Piscataway, NJ. IEEE.

Szewczyk, G., Spinelli, R., Magagnotti, N., Tylek, P., Sowa, J., Rudy, P., and Gaj-Gielarowiec, D. (2020). The mental workload of harvester operators working in steep terrain conditions. *Silva Fennica*, 54(3):10355.

Tiwari, A., Cassani, R., Gagnon, J. F., Lafond, D., Tremblay, S., and Falk, T. H. (2020). Prediction of stress and mental workload during police academy training using ultra-short-term heart rate variability and breathing analysis. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4530–4533. IEEE.

Truschzinski, M., Betella, A., Brunnett, G., and Verschure, P. F. M. J. (2018). Emotional and cognitive influences in air traffic controller tasks: An investigation using a virtual environment? *Applied ergonomics*, 69:1–9.

van Acker, B. B., Bombeke, K., Durnez, W., Parmentier, D. D., Mateus, A. B., Biondi, A., and Vlerick, P. (2020a). Mobile pupillometry in manual assembly: A pilot study exploring the wearability and external validity of a renowned mental workload lab measure. *International Journal of Industrial Ergonomics*, 75:102891.

van Acker, B. B., Conradie, P. D., Vlerick, P., and Saldien, J. (2020b). Employee acceptability of wearable mental workload monitoring: Exploring effects of framing the goal and context in corporate communication. *Cognition, Technology & Work*, 52(1):537–552.

van Acker, B. B., Parmentier, D. D., Conradie, P. D., van Hove, S., Biondi, A., Bombeke, K., Vlerick, P., and Saldien, J. (2020c). Development and validation of a behavioural video coding scheme for detecting mental workload in manual assembly. *Ergonomics*, 64(1):78–102.

van Puyvelde, M., Neyt, X., McGlone, F., and Pattyn, N. (2018). Voice stress analysis: A new framework for voice and effort in human performance. *Frontiers in psychology*, 9:1994.

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics* *and Computing*, 27(5):1413–1432.

Venkatesh, V. and Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46(2):186–204.

Venkatesh, V., Thong, J. Y. L., and Xu, X. (2016). Unified theory of acceptance and use of technology: A synthesis and the road ahead. *Journal of the Association for information Systems*, 17(5):328–376.

Vukovic, M., Sethu, V., Parker, J., Cavedon, L., Lech, M., and Thangarajah, J. (2019). Estimating cognitive load from speech gathered in a complex real-life training exercise. *International Journal of Human-Computer Studies*, 124:116–133.

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., and Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, 25(1):35–57.

Weijters, B., Schillewaert, N., and Geuens, M. (2008). Assessing response styles across modes of data collection. *Journal of the Academy of Marketing Science*, 36(3):409–422.

Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology*, 12(1):3–18.

Xia, F., Yang, L. T., Wang, L., and Vinel, A. (2012). Internet of things. *International Journal of Communication Systems*, 25(9):1101–1102.

Yassaee, M. and Mettler, T. (2019). Digital occupational health systems: What do employees think about it? *Information Systems Frontiers*, 21(4):909–924.

Young, M. S., Brookhuis, K. A., Wickens, C. D., and Hancock, P. A. (2015). State of science: Mental workload in ergonomics. *Ergonomics*, 58(1):1–17.

Zhang, Y., Wang, W., Wu, N., and Qian, C. (2016). IoT-enabled real-time production performance analysis and exception diagnosis model. *IEEE Transactions on Automation Science and Engineering*, 13(3):1318–1332.

Zheng, Y., Lu, Y., Jie, Y., and Fu, S. (2019). Predicting workload experienced in a flight test by measuring workload in a flight simulator. *Aerospace medicine and human performance*, 90(7):618–623.