



## Measuring cognitive workload in the nuclear control room: a review

Per Øivind Braarud

To cite this article: Per Øivind Braarud (26 Jan 2024): Measuring cognitive workload in the nuclear control room: a review, Ergonomics, DOI: [10.1080/00140139.2024.2302381](https://doi.org/10.1080/00140139.2024.2302381)

To link to this article: <https://doi.org/10.1080/00140139.2024.2302381>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 26 Jan 2024.



Submit your article to this journal [↗](#)



Article views: 257




View related articles [↗](#)



View Crossmark data [↗](#)

## Measuring cognitive workload in the nuclear control room: a review

Per Øivind Braarud 

Institute for Energy Technology/OECD, NEA Halden Human Technology-Organisation (HTO) Project, Halden, Norway

### ABSTRACT

Despite the substantial literature and human factors guidance, evaluators report challenges in selecting cognitive workload measures for the evaluation of complex human–technology systems. A review of 32 articles found that self-report measures and secondary tasks were systematically sensitive to human–system interface conditions and correlated with physiological measures. Therefore, including a self-report measure of cognitive workload is recommended when evaluating human–system interfaces. Physiological measures were mainly used in method studies, and future research must demonstrate the utility of these measures for human–system evaluation in complex work settings. However, indexes of physiological measures showed promise for cognitive workload assessment. The review revealed a limited focus on the measurement of excessive cognitive workload, although this is a key topic in nuclear process control. To support human–system evaluation of adequate cognitive workload, future research on behavioural measures may be useful in the identification and analysis of underload and overload.

### PRACTITIONER SUMMARY

This review provides background for the selection of cognitive workload measures for the evaluation of complex human–technology systems and identifies future research needs for applied cognitive workload assessment.

### ARTICLE HISTORY

Received 26 April 2022  
Accepted 2 January 2024



### KEYWORDS

Cognitive workload;  
measurement; control  
room; operators

## 1. Introduction

The nuclear industry provides roughly 10% of the global electricity generation (World Nuclear Association 2022). The control room team plays a vital role in maintaining production, preventing abnormalities, and mitigating potential accidents. High workload is a central topic given the operators' task of interpreting a substantial amount of complex information and, if required, performing critical time-pressured decisions (Woods 1988; Vicente 1999; Ha et al. 2006; Kim, Kim, and Jung 2014; Chen, Yan, and Tran 2019). However, despite the available guidance (O'Hara et al. 2012; Reinerman-Jones et al. 2015; ISO 2016) and substantial research over the last decades (Moray 1988; Young et al. 2015; Charles and Nixon 2019), practitioners of human factors engineering report challenges in selecting cognitive workload measures and interpreting the results of these measures (Pickup, Wilson, and Lowe 2010; Young et al. 2015; OECD NEA 2017; Braarud and Pignoni 2022).

Nuclear control room operators supervise a large industrial system characterised by interactive complexity and tight coupling (Perrow 1984). A wide array of displays and panels are commonly used to oversee and control the process. Given the requirement for reliable and safe production, control room work is guided by operating procedures, communication guidelines, and teamwork guidelines. Concurrent with applying operating procedures, operators monitor the state of the system and anticipate upcoming challenges (Jimmieson and Terry 1999). Consequently, operators may have several simultaneous technical tasks and teamwork activities. Task management activities such as task switching, shedding, and prioritisation are frequently needed (Wickens et al. 2012; Neerincx 2003; Megaw 2005). The choice of operational strategy may create additional or reduced task loads depending on how well the strategy fits the situation (Braarud and Johansson 2010). Investing effort in plant management and the anticipation of challenges may lead to actions

**CONTACT** Per Øivind Braarud  [Per.Oivind.Braarud@ife.no](mailto:Per.Oivind.Braarud@ife.no)  Institute for Energy Technology/OECD, NEA Halden Human Technology-Organisation (HTO) Project, PB 173, NO-1751 Halden, Norway

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

that reduce the future task load, whereas misplaced actions may add problems, causing additional task load. Although the overall workload is within operator limits, peak workload during certain phases of work may exceed the operator's capacity and negatively impact performance (Xie and Salvendy 2000; Gao et al. 2013; Hancock 2017). Given extensive simulator training, the cognitive processing of many generic tasks is substantially automated (Shiffrin and Schneider 1977; Ackerman 1987). Furthermore, operator training includes the management of high task loads. Consequently, control room teams can frequently maintain a given performance level even in highly demanding situations (Bittner 1992; Hockey 1997).

Reducing cognitive workload and maintaining spare capacity in emergencies or stressful situations are the common goals of system design (Vidulich 2000; Wickens 2000; Vidulich and Tsang 2015). Human–system interface (HSI) topics that are frequently related to cognitive workload include the design of alarm systems (Brown, O'Hara, and Higgins 2000; Huang et al. 2006), computer displays (Hwang et al. 2009; Hsieh, Chiu, and Hwang 2014), computerised procedures, and automation (Xu et al. 2008; Lin, Yenn, and Yang 2010). Human Factors validation requires evidence that control room design supports adequate task management and operators safely operate the plant within an acceptable workload envelope (O'Hara et al. 2012; ISO 2017; Simonsen and Osvalder 2018). Validation commonly requires utilising highly demanding scenarios that are performed in control room simulators. Furthermore, cognitive workload is assessed to understand the human performance challenges of the design to gain insights beyond the information provided by the primary task outcome (Parasuraman, Sheridan, and Wickens 2008; Vidulich and Tsang 2015) and evaluate whether the performance observed during testing can be generalised to actual plant operation (De Waard and Evans 2014).

### 1.1. Workload measures

Cognitive workload measures are commonly classified as self-reported, task performance, or physiological (O'Donnell and Eggemeier 1986; Lysaght et al. 1989; Young et al. 2015; Longo et al. 2022). One can also add the class behavioural measures (Parasuraman 2003; Chen et al. 2012; Durantin et al. 2014). Self-report measures require participants to quantify their experience of workload (Tsang and Vidulich 2006). This type of measure is popular given its ease of use and the limited resources needed for implementation (Reid and Nygren 1988). The most popular self-report measure by far is the NASA Task Load Index (NASA-TLX;

Hart and Staveland 1988; Hart 2006; Grier 2015). Self-report measures are frequently collected post-session, i.e. after the scenario or tasks are completed, but can also be applied during task performance (Jordan 1992; Endsley et al. 1998; Carswell et al. 2010). A proposed advantage of self-report techniques is that the operator is aware of the increased cognitive effort that does not necessarily manifest itself in observable performance (Muckler and Seven 1992; Annett 2002). Self-report measures are believed to reflect the number of concurrent tasks and the conscious effort invested by the operators (Gopher and Donchin 1986; O'Donnell and Eggemeier 1986; Yeh and Wickens 1988; Tsang and Vidulich 2006; Cain 2007).

Primary task methods use operator performance as a cognitive workload measure (Tsang and Vidulich 2006). As task demand increases, primary task performance is expected to deteriorate given limited cognitive resources (Yeh and Wickens 1988). However, in complex settings, factors other than cognitive workload may strongly influence primary task performance (Gopher and Donchin 1986; Hancock and Matthews 2019). Secondary tasks measure the remaining operator capacity while primary tasks are performed (Mulder, 1979; Tsang and Vidulich 2006). Commonly used secondary tasks are choice reaction time, time estimation, or memory-search tasks (Wickens et al. 2012).

Physiological approaches measure cognitive workload processes through their effect on the body's state and physiological processes. Electroencephalogram (EEG) techniques measure the electrical activity of the brain through sensors placed on the scalp. Measures focus on the frequency domain and are commonly decomposed into bandwidths (Farmer and Brownson 2003; Charles and Nixon 2019). Electrocardiography (ECG) techniques measure the heart's electrical activity using sensors attached to the chest and limbs. Measurements include the number of heartbeats (per unit of time), the inter-beat interval, heartbeat variability, and frequency-based measures. A head-mounted or remote eye tracker is most often used to measure eye behaviour. Measures include pupil diameter, blink rate, blink duration, and aspects of fixation and saccadic behaviour (Charles and Nixon 2019). An advantage of eye tracker data is that the camera provides the operator's focus area or 'area of interest'. Other physiological measures include respiration rate (Veltman and Gaillard 1998), skin conductance (Or and Duffy 2007; Miyake et al. 2009), and haemodynamic methods that focus on the oxygenation and deoxygenation of the brain's bloodstream (Causse et al. 2017; McKendrick et al. 2019). A main advantage of physiological measures is their ability to generate continuous data (Cain 2007).

Behavioural measures utilise features of operator behaviour. The literature includes broad use of the term. The behavioural basis might range from cursor movements and interface navigation to primary task performance such as response time, accuracy, and task errors (Annett 2002; Parasuraman 2003; Durantin et al. 2014). For the purpose of this review, we will use the term somewhat less broadly than many authors by excluding primary and secondary task measures (Khawaja, Chen, and Marcus 2012; Chen et al. 2012; Braarud and Pignoni 2023). A premise for behavioural measures is that observable overt behaviour may indicate cognitive effort to handle task demand, and that operators adapt their behaviour to manage cognitive workload (Hockey 1997; Hancock and Warm, 1989; De Waard and Evans 2014; Hancock and Matthews 2019). For example, the increased cognitive effort to maintain task performance may be manifested as coping strategies, patterns of interface navigation, information gathering, alarm management, and patterns of team communication and cooperation (Hockey 1997; Hancock and Matthews 2019; Braarud and Pignoni 2023).

Sensitivity, a fundamental measurement criterion, refers to whether the measure discriminates between distinct levels of cognitive workload (O'Donnell and Eggemeier 1986; Wickens et al. 2012). An additional important criterion for complex dynamic work is resolution or granularity (Muckler and Seven 1992; Chen et al. 2012; Chuang et al. 2016). Depending on the purpose of the evaluation, the measurement should be able to provide granularity with regard to work phases or task steps.

## 1.2. Review purpose

Despite human factors guidance and the substantial literature on cognitive workload assessment, researchers, and evaluators of human–technology systems report challenges in selecting and specifying measures. This review supports the specifications of cognitive workload measurement by reviewing the sensitivity of measures to HSI conditions. Furthermore, the review provides an overview of nuclear domain research on cognitive workload methods, evaluates the utility of these methods in the evaluation of HSIs and control room work, and identifies future research needs.

## 2. Method

The review is based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (Moher et al. 2009). Figure 1 illustrates the steps that were involved in the identification and selection of records.

The search terms ('mental workload' OR 'cognitive workload') AND ('nuclear') AND ('control room' OR 'operator') were applied to a full-text search of the Science Direct, PubMed Central, IEEE, and Web of Science repositories. Articles were included if they were published in peer-reviewed journals in English between 1990 and 2020. No ethical approval was needed since this review utilised published peer-reviewed articles only. The search identified 481 records. In total, 404 records were initially excluded because, for example, the empirical work reported on domains other than nuclear or the empirical work reported did not include cognitive workload, e.g. articles that mentioned operator cognitive workload and the nuclear domain only in the Introduction or Discussion. Following this selection, 77 articles were reviewed in full, and 45 were excluded because the full review did not identify a workload measure that was applied, simulation occurred without the involvement of human participants, or the work was purely analytical. Thus, 32 articles were ultimately included for analysis.

The analysis of the 32 articles included recording the cognitive workload measure(s) applied, the purpose of the study, the category of participants, the type of simulation involved, the type of tasks and/or scenarios, and the performance measure(s) applied. The study's purpose was classified as either the evaluation of human–machine interface conditions or the investigation of cognitive workload methods.

The performance measures applied varied substantially across the studies. For example, performance measures that were related to cognitive workload included choosing the wrong object or selecting the wrong command of the procedure step (Choi et al. 2018), the ratio of procedure steps that failed to be finished (Jou et al. 2009; Gao et al. 2013), the alarm detection rate (Lin et al. 2017), and the time needed for judgement (Wu et al. 2016). Although each article's measures may have been useful for their purposes, the variety of performance measures made it difficult to evaluate workload measures based on their relationship with performance across the articles. Consequently, the review was limited to mainly investigating the sensitivity of the workload measures to the factors (e.g. independent variables) that were examined by the given article. In addition, correlations between measures were noted if they were reported in the article.

## 3. Results and discussion

### 3.1. Overview of reviewed papers

The articles covered a total of 11 types of workload measures. The study settings ranged from individual

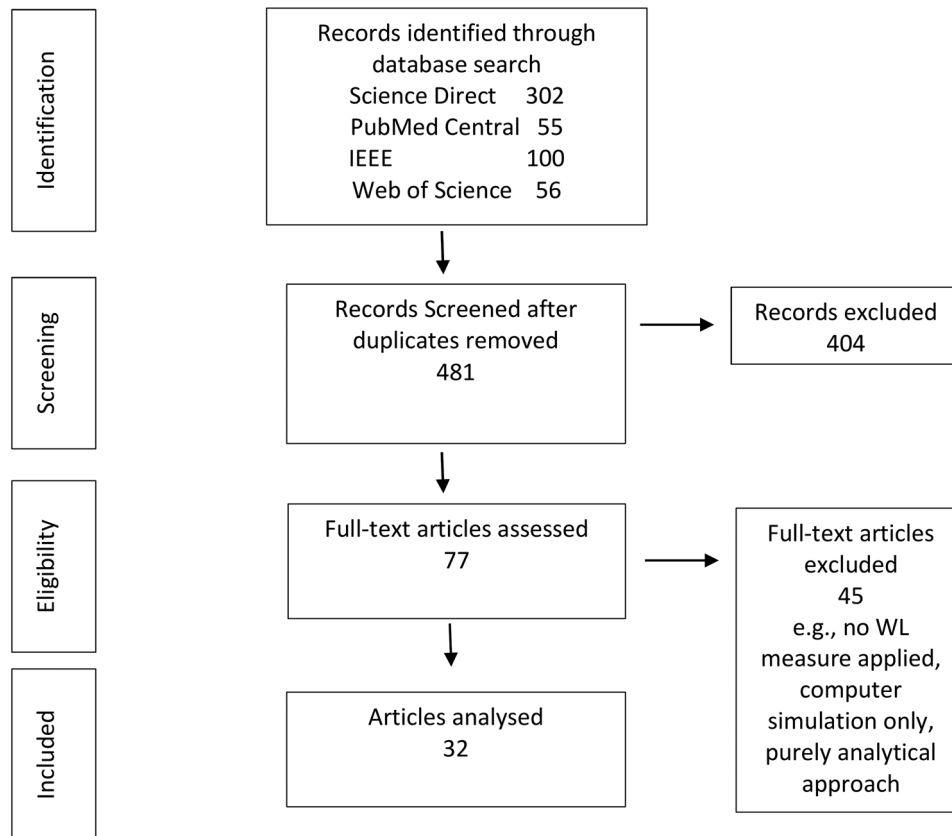


Figure 1. Steps of the literature review.

students performing procedure steps in simplified compact simulations to licenced teams of operators performing demanding accident scenarios in full-scope training simulators. Table 1 provides an overview of the types of measures applied in the articles. A self-report measure alone was the most frequently applied measure (seven articles), followed by a combination of self-report and secondary task measures (five articles). Only four studies did not include a self-report measure. Nine studies included a measure of heartbeats (ECG), eight studies included a secondary task, seven studies utilised operator behaviour, six studies utilised eye behaviour, four studies included measurements of brain wave frequency (EEG), and two studies included primary task performance as measures of workload. Finally, four studies included measurements of haemodynamic processes, speech, skin temperature, or respiration.

Eighteen of the 32 articles addressed the development and empirical evaluation of workload measures. The remaining 14 articles utilised workload measures in the evaluation of HSIs. Table 2 provides an overview of the topics of the articles that were related to cognitive workload measurement. Alarm systems were the most frequently studied HSI topic.

Ten articles used teams of licenced nuclear control room operators in full-scale simulators (Braarud 2020; Braarud et al. 2020; Chuang et al. 2016; Chung, Yoon, and Min 2009; Gan et al. 2020; Kim, Kim, and Jung 2014; Lau et al. 2008; Lin et al. 2017; Park et al. 2017; Park, Jung, and Kim 2020). The majority (22) of the articles used student participants in compact simulators. In four of these studies, the participating students were organised in teams consisting of control room roles (Hwang et al. 2009; Lin, Hsieh, and Lin 2013; Reinerman-Jones, Matthews, and Mercado 2016; Yang et al. 2012). The remaining 18 articles described individual students (Al Harbi et al. 2013; Chen, Yan, and Tran 2019; Choi et al. 2018; Gao et al. 2013; Ha et al. 2006; Hsieh et al. 2012; Hsieh, Chiu, and Hwang 2014; Hsieh, Chiu, and Hwang 2015; Huang et al. 2006; Hwang, Lin, et al. 2008; Hwang, Yau, et al. 2008; Jou et al. 2009; Lin, Yenn, and Yang 2010; Reinerman et al. 2020; Wu et al. 2016; Wu et al. 2020; Xu et al. 2008; Yan et al. 2017). In most cases, the students were engineering students; some of these students were nuclear engineering students. Whereas studies that described operators utilised scenarios, studies of students commonly investigated the performance of a given operating procedure. The

**Table 1.** Type of cognitive workload measure applied by the articles.

Self-report	Primary task	Secondary Task	Behavioural	EEG	Eye	ECG	Skin	Speech	Respiration	Hemo dynamic	Reference	Number
✓											Braarud 2020; Hsieh et al. 2012; Hsieh, Chiu, and Hwang 2014; Huang et al. 2006; Lau et al. 2008; X. Wu et al. 2016; Xu et al. 2008	7
✓		✓									Hsieh, Chiu, and Hwang 2015; Hwang, Lin, et al. 2008; Lin, Yenn, and Yang 2010; Lin, Hsieh, and Lin 2013; Yang et al. 2012	5
✓					✓						Ha et al. 2006; Wu et al. 2020; Yan et al. 2017	3
✓			✓								Chuang et al. 2016; Park et al. 2017; Park, Jung, and Kim 2020;	3
			✓								Chung, Yoon, and Min 2009; Kim, Kim, and Jung 2014	2
✓	✓				✓						Chen, Yan, and Tran 2019	1
✓					✓	✓					Gao et al. 2013	1
✓		✓			✓	✓					Hwang, Yau, et al. 2008	1
✓			✓			✓			✓		Gan et al. 2020	1
✓						✓					Hwang et al. 2009	1
✓	✓	✓				✓					Jou et al. 2009	1
✓		✓				✓					Lin et al. 2017	1
✓				✓		✓				✓	Reinerman-Jones, Matthews, and Mercado 2016	1
✓				✓		✓					Reinerman et al. 2020	1
✓			✓						✓		Braarud et al. 2020	1
✓				✓		✓					Choi et al. 2018	1
				✓		✓	✓	✓			Al Harbi et al. 2013	1
28	2	8	7	4	6	9	1	1	1	1	Number of articles	32

studies of students included training on the simulation and information on the specific operating procedures under investigation.

### 3.2. Self-report workload measures

Twenty-four of the 28 articles that addressed self-report measures applied the NASA-TLX (Hart and Staveland 1988). In one of these articles, Ha et al. (2006) applied the Modified Cooper Harper (MCH) scale (Wierwille and Casali 1983) and the NASA-TLX. Braarud et al. (2020) applied an additional self-report measure developed specifically for the study. Two additional articles used the MCH scale (Park et al. 2017; Park, Jung, and Kim 2020) and Lau et al. (2008) used the Halden Task Complexity Questionnaire (Braarud 2000). Reinerman et al. (2020) applied the Multiple Resource Questionnaire (Boles and Adair 2001) and the Instantaneous Self-Assessment of Workload technique (Jordan 1992). Regarding HSI studies, Hsieh et al. (2012) reported significantly lower NASA-TLX when an alarm procedure support system was available compared with no support. NASA-TLX ratings were also significantly lower with pre-alarm support compared with no pre-alarm support (Hwang, Lin, et al. 2008; Lin et al. 2017). These results corresponded with the reported significant sensitivity of a secondary task. However, Wu et al. (2016) found no significant effect on the NASA-TLX of bar

versus tile alarm presentation, and Huang et al. (2006) reported no significant effect of manual versus automatic alarm reset. Yang et al. (2012) observed that supervisors rated the NASA-TLX workload significantly lower when they used computerised procedures compared with paper-based procedures. A similar significant effect was found for a secondary task. By evaluating display design, Hsieh, Chiu, and Hwang (2015) found a significant effect of the quantity of display information on the NASA-TLX. A similar significant effect was found for a secondary task. Yan et al. (2017) reported lower NASA-TLX scores for displays based on human factors engineering principles compared with the original design. However, two studies that compared displays based on ecological design principles with traditional displays did not find significant main effects on self-reported workload (Hsieh, Chiu, and Hwang 2014; Lau et al. 2008). A hypothetical explanation of this finding is that studies that manipulated the presentation format did not affect participants' deliberate cognitive effort to the same extent as those that manipulated task content. Several studies reported significantly lower self-reported workloads for higher levels of automation than for lower levels of automation (Hsieh, Chiu, and Hwang 2014; Jou et al. 2009; Lin, Yenn, and Yang 2010). Lin, Yenn, and Yang (2010) reported corresponding effects on a secondary task, whereas Jou et al. (2009) reported no significant effect



**Table 2.** Overview of the articles' topic.

Topic	Reference	Number
Alarms, attention support	Hsieh et al. 2012; Huang et al. 2006; Hwang, Lin, et al. 2008; Lin et al. 2017; Wu et al. 2016	5
Computerised procedures	Hsieh, Chiu, and Hwang 2015; Xu et al. 2008; Yang et al. 2012	3
Displays/System control	Hsieh, Chiu, and Hwang 2014; Lau et al. 2008; Yan et al. 2017	3
Automation	Jou et al. 2009; C.J. Lin, Yenn, and Yang 2010	2
Soft controls	Al Harbi et al. 2013	1
Cognitive workload method	Braarud 2020; Braarud et al. 2020; Chen, Yan, and Tran 2019; Choi et al. 2018; Chuang et al. 2016; Chung, Yoon, and Min 2009; Gan et al. 2020; Gao et al. 2013; Ha et al. 2006; Hwang, Yau, et al. 2008; Hwang et al. 2009; Kim, Kim, and Jung 2014; Lin, Hsieh, and Lin 2013; Park et al. 2017; Park, Jung, and Kim 2020; Reinerman-Jones, Matthews, and Mercado 2016; Reinerman et al. 2020; Wu et al. 2020	18

of the automation level on a secondary task. Braarud (2020) provided an example of the cognitive workload evaluation of an integrated control room; by comparing the operator's workload rating of the modernised control room to the rating of the old control room, the author reported a significant and relatively large effect on NASA-TLX mental demand.

Studies that assessed competence and experience levels found that self-reported workload was significantly lower for experienced participants compared with less experienced participants (Park et al. 2017; Park, Jung, and Kim 2020; Wu et al. 2020). Xu et al. (2008) reported that participants rated NASA-TLX significantly lower after 15 trials with an operating procedure than after the first five trials. Regarding staffing, Lin, Hsieh, and Lin (2013) found that the cognitive workload of a one-person team was significantly higher than that of a two-person team. A similar effect was observed for a secondary task. Yang et al. (2012) found that people in different positions within a team, i.e. supervisor, reactor operator, and assistant reactor operator, rated NASA-TLX significantly differently, whereas Hwang, Lin, et al. (2008) observed no significant effect of the operator role on a two-person team. However, both studies reported a significant effect of position on a secondary task. The sensitivity of self-report measures to team position has been supported by other studies (Hill et al. 1989; Braarud 2021).

Five HSI studies analysed the effect of tasks or scenarios. Four studies reported significant effects (Hwang, Lin, et al. 2008; Lau et al. 2008; Hsieh, Chiu, and Hwang 2015; Lin et al. 2017), whereas one study (Hsieh, Chiu,

and Hwang 2014) reported that the NASA-TLX was not significantly sensitive. Most method studies investigated the effect of task or scenario complexity and found that self-reported workload measures were sensitive to levels of task or scenario complexity (Ha et al. 2006; Gao et al. 2013; Chuang et al. 2016; Reinerman-Jones, Matthews, and Mercado 2016; Braarud 2020). For example, Ha et al. (2006) utilised eight accident diagnosis tasks of substantially varying demand and reported correlations of  $r = .89$  and  $r = .91$  between the task demand and NASA-TLX and MCH, respectively. Braarud (2020) reported that scenarios that ranged from procedure-guided operation to challenging knowledge-based tasks explained 26.3% of the variance in operator TLX mental demand. In the method studies, the reported effects on self-reported cognitive workload frequently co-occurred with the same direction effects on other types of workload measures. For example, Choi et al. (2018) reported a correlation of .84 for the NASA-TLX and an EEG-based index, Chent et al. (2019) reported a correlation of .46 between pupil size and TLX, and Ha et al. (2006) reported significant correlations between TLX or MCH and several ocular measures that ranged from .66 to .93. The reported sensitivity to task and scenario complexity corresponds with a recent review of the NASA-TLX (Hertzum 2021) and studies in aviation (Vidulich and Tsang, 1986; Battiste and Bortolussi 1988; Corwin et al. 1989).

The results reported for the frequently applied NASA-TLX plausibly extend to other self-report measures. For example, Ha et al. (2006) reported a significant correlation of .89 between NASA-TLX and MCH. Braarud et al. (2020) reported a significant correlation of .55 between the NASA-TLX and a study-specific self-report measure for control room work. Correspondingly, studies in other domains report high correspondence between self-report measures such as NASA-TLX, MCH, and SWAT (Vidulich and Tsang 1986; Hill et al. 1992; Rubio et al. 2004).

### 3.3. Task measures

#### 3.3.1. Primary task measures

Whereas most articles included measurements of primary task performance, two articles utilised primary task performance as a workload measure. In their method study, Chen, Yan, and Tran (2019) reported that time spent on a task correlated negatively with the blink rate ( $r = -.54$ ) and that the error rate correlated positively with the fixation rate ( $r = .45$ ). Jou et al. (2009) reported that for the task of reactor shutdown, participants were in

the low automation condition for a significantly longer time compared with the high automation condition. Although extensively used in highly controlled settings (O'Donnell and Eggemeier 1986), the limited application of primary task workload measures may reflect that in process control, many factors in addition to cognitive workload influence primary task performance (O'Hara et al. 2012; Wu and Li 2013). Furthermore, the definition of primary tasks, i.e. time spent on an operating procedure step or an error in such as step, may not be considered primary task performance according to the functional goals of supervising and controlling a nuclear power plant (O'Hara et al. 2012).

### 3.3.2. Secondary task measures

Secondary measures were mainly applied for HSI evaluation and were not investigated by method studies. The majority of studies that investigated secondary tasks utilised randomly presented visual choice reaction time tasks (Lysaght et al. 1989) such as deciding between single versus double dots or on a dot colour, or comparing integers (Hsieh, Chiu, and Hwang 2015; Lin, Yenn, and Yang 2010; Lin, Hsieh, and Lin 2013; Hwang, Lin, et al. 2008; Lin et al. 2017). The recorded performance included ratios of hits, misses false alarms, correct rejections, and response time. Two articles investigated mental arithmetic tasks (Yang et al. 2012; Hwang, Yau, et al. 2008).

All of the articles that applied secondary measures also applied a self-report measure. Therefore, the main portion of the results is included in the section above that addresses self-report measures (Hwang, Lin, et al. 2008; Lin et al. 2017; Yang et al. 2012; Hsieh, Chiu, and Hwang 2015; Lin, Yenn, and Yang 2010) and is not repeated here. In summary, secondary tasks were sensitive to similar HSI conditions as described for the self-report measures, including alarm system design (Hwang, Lin, et al. 2008; Lin et al. 2017), computerised procedures (Yang et al. 2012), display design related to information quantity and design principles (Hsieh, Chiu, and Hwang 2015; Yan et al. 2017), level of automation (Hsieh, Chiu, and Hwang 2014; Jou et al. 2009; Lin, Yenn, and Yang 2010), and staffing and team composition (Lin, Hsieh, and Lin 2013; Yang et al. 2012; Hwang, Lin, et al. 2008). The results correspond with findings on automation levels, equipment design, and operator support within aviation and air-traffic management (Slocum, Williges, and Roscoe 1971; Perry, Segall, and Kaber 2005; Helmke et al. 2016), and scenarios (Bortolussi, Kantowitz, and Hart 1986; Bortolussi, Hart, and Shively 1987). However, the results for secondary tasks may also depend on the performance characteristics used. For example,

Lin, Yenn, and Yang (2010) found no significant effect of the level of automation on the correctness of the secondary task response; however, the secondary task response time was significantly affected.

### 3.4. Behavioural measures

Seven methodological articles investigated cognitive workload measures that can be labelled 'behavioural'. Park et al. (2017) reported a significantly higher number of team interface management tasks in complex accident scenarios than in less complex scenarios, and Park, Jung, and Kim (2020) reported a significant correlation between self-reported workload (i.e. using the MCH) and the number of interface management tasks in accident scenarios. Braarud et al. (2020) observed that interface management tasks predicted the operators' cognitive workload for two-minute segments of scenarios with an accuracy of .61. The operator's acknowledgement of alarms was an important feature for predicting the operator's cognitive workload. Kim, Kim, and Jung (2014) reported descriptive results that supported a substantial difference in the frequency of interface management tasks between control room positions. These results correspond with research in other domains that reports correlations between cognitive workload and interface management activities (Chen et al. 2012; Arshad, Wang, and Chen 2013; Lin, Hsieh, and Lin 2013; Tobaruela et al. 2014; Pimenta et al. 2016). However, behavioural methods extend beyond the classification of interface activity. In addition to reporting on interface management, Kim, Kim, and Jung (2014) observed that the frequencies of cognitive and communicative activities were higher for supervisors than for those in other control room positions. Chung, Yoon, and Min (2009) recorded the frequency of teams' communication threads and proposed that the number of simultaneous threads, type of message, and timing, e.g. delayed response, could be workload indicators. Chuang et al. (2016) related approach classified operator behaviour according to Rasmussen's (1986) skill-, rule-, and knowledge-based categories. The author found that the two ratios that involved rule- and knowledge-based behaviour over total behaviour were correlated with the NASA-TLX ( $r = .51$  and  $r = .56$ , respectively). Finally, Gan et al. (2020) estimated workload scores for operator activities (McCracken and Aldrich 1984) based on video recordings of operator performance. The authors reported significant correlations between behavioural-based estimates and heart rate variability ( $\rho = -.42$ ) and breathing wave amplitude ( $\rho = -.46$ ).



### 3.5. Physiological measures

#### 3.5.1. Electroencephalogram

Four studies utilised EEG measures. Three of these were method studies, whereas one article evaluated soft controls. Al Harbi et al. (2013) reported that the beta power ratio was higher when soft versus hard controls were used; however, the difference was not statistically significant. The increased beta power ratio was assumed to represent increased alertness. Reinerman-Jones, Matthews, and Mercado (2016) observed that theta, beta, and gamma were sensitive to the type of task, suggesting a higher cognitive workload for checking and response implementation tasks than for detection tasks. However, participants' NASA-TLX rating was significantly higher for detection tasks than for checking and response implementation, suggesting some uncertainty about the interpretation of the sensitivity of EEG measures. A decrease in alpha band power and an increase in frontal theta have generally been related to increased task demand, and several EEG measures have been reported to be sensitive to different tasks (Borghini et al. 2014; Charles and Nixon 2019). A plausible interpretation is that the HSI controls and types of tasks studied did not represent sufficient variation in task demand to elicit the expected effects on the EEG measures applied. Studies have also found patterns in EEG results in the repetition of tasks. Reinerman et al. (2020) reported examples of individual participants' linear, quadratic, and cubic effects across 27 sessions for alpha, beta, and theta waves. Finally, Choi et al. (2018) utilised alpha, beta, theta, and gamma powers in their development of an EEG-based workload index and reported a correlation of .84 between the index and the NASA-TLX. This article is further described in Section 3.6 on indexes.

#### 3.5.2. Ocular measures

Six studies utilised ocular measures. Five of these were method studies, whereas one article evaluated system displays (Yan et al. 2017). Yan et al. (2017) reported a significantly lower blink rate, higher fixation duration, and higher fixation rate for an original design compared with a human factors-designed interface. Pupil dilation did not differ significantly. Furthermore, NASA-TLX scores were significantly higher for the original interface compared with the redesigned interface – a workload indication corresponding with the ocular measures. The methodological studies found that the blink rate decreased with increased self-reported workload (Ha et al. 2006; Chen, Yan, and Tran 2019; Wu et al. 2020). Wu et al. observed that non-experts showed a significantly lower blink rate than experts,

providing additional evidence that a lower blink rate indicates a higher cognitive workload. The results correspond with the literature, which reports a decreased blink rate due to increased visual demand (Charles and Nixon 2019). However, Gao et al. (2013) reported opposing results, in which a higher blink rate was related to an increased task load. Ha et al. (2006) observed that an increased task load was strongly associated with an increased number of fixations and an increased fixation. The authors further reported significant positive correlations between the number of fixations and self-reported workload and between fixation duration and self-reported workload. In contrast, fixation duration in aviation has been reported to decrease with increasing task load during flights (De Rivecourt et al. 2008). These conflicting findings illustrate the challenges of interpreting fixation results across tasks and domains. The increased fixation duration of the nuclear control room tasks may indicate increasing difficulty in interpreting the information presented, whereas the decreased fixation duration of an increased task load in flights may indicate that more information must be addressed more frequently, thereby leading to a decreased fixation duration. Furthermore, Wu et al. (2020) reported a significant positive correlation between fixation rate and self-reported workload for experts, and Chen, Yan, and Tran (2019) found that neither fixation duration nor fixation rate correlated significantly with self-reported workload. Chen, Yan, and Tran (2019) and Wu et al. (2020) reported a non-significant increase in pupil size was related to increased self-reported workload. In addition, Gao et al. (2013) reported a trend of larger pupil size for high scenario complexity compared with low scenario complexity. Pupil dilation appears to be a stable physiological feature; this corresponds with the literature (Charles and Nixon 2019); however, the interpretation fixations may depend on the task context. One study (Hwang, Yau, et al. 2008) included ocular measures in its development of performance models but did not specifically report results for each ocular measure. The model results are included in Section 3.6 on workload indexes.

#### 3.5.3. Electrocardiographic activity

Nine studies utilised various ECG measures for the assessment of cognitive workload. Six of the nine articles addressed workload method development, whereas three articles reported studies of HSI. Al Harbi et al. (2013) reported that high-frequency parasympathetic activity was lower for soft controls (indicating a higher workload) than for hard controls, although the

result was not statistically significant. The remaining two studies that evaluated human–system factors observed no significant effects on the average heart rate of scenarios, level of automation (Jou et al. 2009), or pre-alarm support (Lin et al. 2017). As suggested in the literature, the lack of sensitivity may result from the fact that differences in task demand between HSI conditions must be high to be reflected in cardiac activity, especially for heart rate variability (HRV; Mulder et al. 2000; Charles and Nixon 2019). However, the conditions studied by Jou et al. (2009) and Lin et al. (2017) differed sufficiently to reveal an effect on the operators' NASA-TLX rating.

In the articles that addressed workload method, the studies revealed significant effects on heart rate and HRV from task step demand (Gan et al. 2020), scenario complexity (Gao et al. 2013), and type of task (Reinerman-Jones, Matthews, and Mercado 2016). Gan et al. (2020) showed that heart rate and HRV were related to analytically estimated task demand (McCracken and Aldrich 1984) across an accident scenario of 120 task steps divided among four team members ( $\rho = .40$  and  $\rho = -0.42$ , respectively). Gao et al. (2013) reported higher HRV for a complex scenario than for a non-complex scenario; this co-occurred with an increased blink rate. The finding of increased HRV is contrary to the general finding of decreasing HRV when cognitive demand increases and may be associated with the longer duration of the high-complexity scenario compared with the low-complexity scenario (Mulder and Mulder 1981; Charles and Nixon 2019). Two studies that examined task repetition showed mixed results. Reinerman-Jones, Matthews, and Mercado (2016) reported that the mean HRV increased across four repetitions of detection steps; no such effect was observed for heart rate. However, in their detailed investigation of three participants, Reinerman et al. (2020) reported no statistically significant trends in HRV across 27 sessions and three task types. However, a significant cubic effect on heart rate was identified in one of the three participants for one of the three tasks. Significant effects on HRV were identified for two of the three participants for one of the three task types. Hwang, Yau, et al. (2008) Hwang et al. (2009) developed models for predicting performance. Hwang, Yau, et al. (2008) reported that HRV, heart rate, and systolic pressure were significant indicators of performance, whereas the low frequency/high frequency ratio or diastolic pressure was not an indicator. Finally, Hwang et al. (2009) classified the team's performance with full accuracy by using a model of eight ECG indicators. However, the model was validated with data

from only three teams (a total of three classifications; see Section 3.6 on indexes).

#### 3.5.4. Skin temperature, speech features, respiration, and hemodynamic indicators

One study utilised skin temperature, speech features, respiration, or hemodynamic indicators for the measurement of cognitive workload. In their study of HSI, Al Harbi et al. (2013) observed a higher number of participants who showed a drop in skin temperature – an indicator of stress – when using soft controls compared with those who used hard controls. No significance test was presented for this finding, but the results corresponded with poorer workload indicators from ECG and EEG for the soft-control condition, and the skin temperature correlated highly with the procedural error rate ( $R^2 = .88$ ). Braarud et al. (2020) observed that speech features that were extracted from control room team communication predicted operator cognitive workload for two-minute segments of the scenario with an accuracy of .63. The most important speech variables were the fundamental frequency (pitch), articulation rate, and amplitude. Gan et al. (2020) found that respiration rate and analytically estimated task demand were correlated ( $\rho = .27$ ). Moreover, breathing wave amplitude was correlated with the estimated workload ( $\rho = -.46$ ). A noteworthy aspect of this study was that the respiration rate was collected from operators in a four-person team that operated in a full-scale simulator. Reinerman-Jones, Matthews, and Mercado (2016) utilised hemodynamic measurements. The authors reported that right hemisphere cerebral blood flow velocity was sensitive to the task type and that left and right hemisphere oxygen saturation based on functional near-infrared spectroscopy measurements were sensitive to the task type.

#### 3.6. Indexes of workload measures

Seven articles focused on the development of a workload index based on several measures. All of these articles reported methodological studies. Where reported by the authors, the sensitivity of individual measures used in these studies is presented in the respective sections above. The review first examined articles that reported an index from one type of measure. Choi et al. (2018) used the EEG measures theta, alpha, beta, and gamma. The developed index was validated against the NASA-TLX that was collected post-session. The authors reported a .84 correlation between the index and NASA-TLX and a correlation of .83 between the index and procedure errors. Wu et al. (2020) reported that an

index developed from ocular measures (pupil diameter, blink rate, fixation rate, and saccadic rate) predicted the NASA-TLX score with an accuracy of  $R = .98$ . The NASA-TLX ratings were collected post-session. Hwang et al. (2009) utilised several ECG measures (heart rate, HRV, and HRV frequencies) to develop an algorithm to predict a performance index based on correct response and response time for procedure steps. The performance used in the prediction was aggregated to one score per participant. Algorithm evaluation using data from three teams (three predictions performed) showed that the algorithm classified the correct performance class out of three classes, which provided an accuracy of 1. Gao et al. (2013) constructed a model from ocular (pupil size, blink rate, and blink duration) and ECG (HRV and HRV frequency) measures. The model was evaluated with data from two participants and predicted post-session NASA-TLX with an accuracy of  $R^2 = .77$ . No individual measures correlated significantly with the NASA-TLX; therefore, the index results suggested an improvement over the individual measures. Hwang, Yau, et al. (2008) developed a model using ocular (blink rate and blink duration), ECG (heart rate, HRV, and HRV frequency), and blood pressure (systolic and diastolic) measures. The model was evaluated by predicting the error rate on a secondary task for two participants, resulting in  $R^2 = .84$ . Chen, Yan, and Tran (2019) developed an index from a broad set of data that consisted of the NASA-TLX, performance indicators, and ocular measures (pupil dilation, blink rate, fixation duration, and fixation rate). The index was significantly sensitive to the interface type. Braarud et al. (2020) developed a model from operator speech features and human-system interaction activities. A model trained on team member data predicted operator workload for two-minute segments of the scenario with a classification accuracy of .72. Corresponding with the index results, relatively high accuracy of workload indexes in the prediction and classification of cognitive workload has been reported in other settings (Wilson and Russell 2003; Chen et al. 2012; Solovey et al. 2014; Borghini et al. 2014). However, examples of workload indexes that provide only moderate accuracy in other settings are also available (Charfuelan and Kruijff 2013; McDonald, Ferris, and Wiener 2020).

#### 4. Considerations for the evaluation of control room HSI

A remarkable characteristic of the articles was the frequent use of self-report measures for the evaluation of HSI and their frequent use as a criterion for the investigation of physiological measures. Self-report

measures were sensitive to alarm procedure selection, pre-alarm design, computerised procedures, the level of automation, and display information quantity. Self-report measures also correlated with secondary tasks and physiological measures. Self-reported workload measures were sensitive to conditions that affected operator task content and conditions that explicitly involved resources specific to expertise and staffing. This suggests that self-report measures are suitable for evaluating HSI, task design, and the organisation of work – especially if the work involves deliberate effort, judgement, and decision-making (Gopher and Donchin 1986; O'Donnell and Eggemeier 1986; Yeh and Wickens 1988; Tsang and Vidulich 2006; Cain 2007). Consequently, a self-reported cognitive workload measure should be included in the evaluation of control room work and HSI. However, the granularity that is required for measurement should be considered (Muckler and Seven 1992; Chen et al. 2012; Chuang et al. 2016). Plausibly, a post-scenario assessment may be sufficient for short performance episodes such as a section of an operating procedure or a few procedure steps. For longer performance episodes including realistic event and accident scenarios, utilising the online rating of a single item measure such as the Instantaneous Self-Assessment of Workload (Jordan 1992) technique or the Rating Scale of Mental Effort (Zijlstra 1993) may be considered. Alternatively, the NASA-TLX item demand or the item effort may be used given that these items closely resemble the average of all six NASA-TLX items (Hertzum 2021), particularly for cognitive control room work (Braarud 2020).

Primary task measure has limited utility in complex settings (Wu and Li 2013; Hancock and Matthews 2019); two articles utilised only such measures. However, the results suggest that secondary tasks provide sensitivity to HSI evaluation similar to self-report measures. The extent to which secondary tasks reflect spare resources for the primary task or resources that are involved in task management has been discussed (Lysaght et al. 1989). However, task management is an essential aspect of complex settings.

No study of HSI utilised behavioural measures for cognitive workload evaluation. However, this approach has been used in other domains (Chen et al. 2012; Pimenta et al. 2016; McDonald, Ferris, and Wiener 2020) and may require further development for HSI evaluation in the nuclear domain (Braarud and Pignoni 2023).

Few articles reported the use of physiological workload measures for HSI evaluation. One study that utilised EEG reported non-significant sensitivity, whereas one study that used ocular measures reported significant sensitivity. None of the three studies that utilised

ECG measures reported significant sensitivity, and one study that used skin temperature reported significant sensitivity. One study each of skin temperature, speech features, respiration, and haemodynamic indicators provided promising results, but these measures require further investigation regarding control room HSI evaluation. Method study results regarding EEG, ocular, and ECG measures were mixed. In summary, evidence from the reviewed articles was insufficient to support a clear recommendation of physiological measures for the evaluation of HSI. A similar conclusion has been previously proposed by Farmer and Brownson (2003) for air-traffic management and for general use by Cain (2007). Furthermore, a recent review of physiological workload measures concluded the current absence of a basis for recommending any single physiological measure for the assessment of cognitive workload (Charles and Nixon 2019). However, ocular workload measures based on eye tracker data have a notable advantage for the evaluation of work that involves visual information-gathering across several surfaces. Measuring areas of interest or utilising the scene camera directly provides information particularly useful for the analysis of factors that impact the operator's cognitive workload. Furthermore, the reviewed method studies indicated that blink rate decreased with increased task load (Ha et al. 2006; Chen, Yan, and Tran 2019; Wu et al. 2020) and pupil size increased with increased workload (Chen, Yan, and Tran 2019; Wu et al. 2020). These findings correspond with reports in the literature (Charles and Nixon 2019).

However, the review showed positive evidence for physiological index measures. In summary, the articles that addressed workload indexes reported a high correlation with the NASA-TLX (Gao et al. 2013; Choi et al. 2018; Wu et al. 2020), a high correlation with a secondary task measure (Hwang et al. 2009), perfect accuracy in predicting operator workload (Hwang et al. 2009), and moderate accuracy in classifying operator workload (Braarud et al. 2020). In addition, one study showed that the index was sensitive to the interface type (Chen, Yan, and Tran 2019). However, these indexes are associated with higher cost and complexity compared with self-report measures (Farmer and Brownson 2003; Chuang et al. 2016).

The evaluation of an acceptable or optimal cognitive workload is an important and challenging topic for control room work. This challenge relates to the interpretation of measurement results. Rating scales have the advantage of having a standardised scale bound by endpoints that apply across participants and situations. The labelling of the scale facilitates the interpretation of the scores. Although the actual

understanding of the labels may vary between people, the labelling supports the interpretation of the ratings. The definition of the scale and its labelling may also explicitly address the acceptability of the workload (Jordan 1992; Colle and Reid 2005; Braarud 2020). Secondary tasks are sensitive to high and excessive task demands (Lysaght et al. 1989) and may be suitable for the development of test scores that indicate the level of cognitive workload. Secondary tasks have the advantage of providing a common metric that can be used for comparison across tasks and settings (Wickens et al. 2012). In this respect, a challenge for physiological measures is that the measurement value itself is difficult to interpret for an adequate level of cognitive workload. As such, this is an important additional reason against recommending current physiological measures for the evaluation of nuclear control room HSI.

In most complex work settings, measurement, e.g. self-reported cognitive workload or physiological indicators, is usually insufficient for understanding how HSI or work organisation affects operator cognitive workload. Performing debriefs after a scenario is therefore recommended for the further investigation of cognitive workload in the context of specific task demands and HSIs (Lysaght et al. 1989; Braarud and Pignoni 2022).

## 5. Future research needs

Based on the positive evidence for self-report measures collected post-session, a similar quality of continuous or periodic self-report during work can be suggested. However, adequate methods and the validity of continuous self-reporting of cognitive workload should be investigated. In addition, research on physiological measures can benefit from improved self-report measures.

The reviewed articles investigated different secondary tasks. A standardised secondary task can provide a common metric that can be used to compare workload across tasks and settings (Wickens et al. 2012). Future research can investigate one or a few adequate and general secondary tasks that apply to nuclear control room evaluation. Furthermore, Lysaght et al. (1989) concluded that embedded secondary tasks which are unique for each human-machine setting are useful for complex system evaluation. No article explored embedded secondary task performance. Future research could investigate embedded secondary tasks for cognitive workload measurement related to operator utilisation of the control room HSI.

The method studies on behavioural measures reported counts and classifications of operator activity



that were related to self-reported workload (Chuang et al. 2016; Park, Jung, and Kim 2020; Braarud et al. 2020) and were sensitive to scenario complexity (Park et al. 2017) and control room position (Kim, Kim, and Jung 2014). Behavioural measures may be especially useful for detecting high cognitive workload based on the assumption that operators who experience a high cognitive workload exhibit observable behaviour related to the adaptation to and management of the workload (Hockey 1997; Chen et al. 2012; Braarud and Pignoni 2023). Similar to embedded secondary tasks, the activity measures are nonintrusive and thus do not affect operator tasks and can easily be applied to nuclear control room work. Furthermore, behavioural measures can be applied without resource-demanding dedicated data registration equipment (Chen et al. 2012). However, the reviewed approaches require further development. For example, the extent to which the behavioural indicators and activity classifications truly indicate cognitive workload (Kim, Kim, and Jung 2014) and the extent to which the measures indicate the proportion of cognitive capacity expended can be investigated.

The review provided examples of promising methodological studies on physiological cognitive workload measures for complex work settings. Charles and Nixon (2019) identified a growing empirical basis that can inform future studies about physiological workload measures. However, for the evaluation of HSI in complex work settings, future research must investigate the utility and validity of physiological measures relative to the utility and validity of self-report and secondary task methods. Beyond providing continuous measurement, physiological measures have a role in evaluating cognitive workload that is not accessible to operator self-reporting and is not captured by secondary task approaches.

The results reported indicate the potential for index-derived workload measures. To date, the literature includes relatively few studies within the area of process control compared with studies in domains such as aviation and driving (Borghini et al. 2014). However, the results of the accuracy and utility of the indexes were insufficient in detail because the studies frequently compared aggregated index results with post-session self-reports. Future research must demonstrate the utility of the indexes in HSI evaluation.

Several articles motivate research on cognitive workload by describing the negative consequences of high workload on operator performance. The ultimate concern of cognitive workload measures for human-system evaluation is the adequacy of the interpretations and decisions regarding HSI, operator competence,

and operator performance that can be based on the measurement (Messick 1990). However, the focus on the measurement of excessive cognitive workload is limited. Theory suggests the qualitatively different zones of workload (O'Donnell and Eggemeier 1986; Rasmussen 1986; Young et al. 2015) and measures that can support the classification of underload or overload can be useful in human-system design and evaluation. Future research that supports the analysis and classification of cognitive workload by behavioural indicators may improve the utility of cognitive workload measurement for human-system design and evaluation.

Finally, several authors have called for the future investigation of cognitive workload utilising realistic control room simulation and control room tasks, and teams of experienced operators (Xu et al. 2008; Hsieh et al. 2012; Wu et al. 2016; Lin et al. 2017; Yan et al. 2017).

## 6. Limitations

The articles that were eligible for the review provided an unbalanced representation of measurement types. For example, self-report measures were highly overrepresented. However, the overall interpretations and conclusions about cognitive workload measures for HSI evaluation do not conflict with the literature on the less frequently applied measures addressed in the reviewed papers. The studies reviewed varied substantially in the performance measures, tasks, and scenarios investigated – a phenomenon that is not uncommon in complex domains. Therefore, evaluating workload measures across the articles based on their correspondence with performance results was not meaningful. Consequently, the review was limited to investigating the sensitivity of workload measures. If reported in articles, correlations between cognitive measures were included. Therefore, the literature should be consulted regarding aspects of validity not covered by this review. No grey literature or conference papers were included in the review. The inclusion of only peer-reviewed journal articles assured the quality of the included articles. However, important findings and proposals for new methods may be available only in the grey literature. Future reviews could include a broader range of literature types.

## 7. Conclusion

Self-report measures and secondary task measures were systematically sensitive to human-machine conditions in the nuclear control room and correlated with physiological measures. This finding presumably



extends to similar complex work settings. Including a self-report measure of cognitive workload or secondary task measures in the evaluation of HSIs is therefore recommended. Physiological measures may have the potential to capture cognitive workload that is not accessible for self-reporting and can thereby provide dynamic measurement. However, research must demonstrate the utility of physiological measures in process control settings and human–system evaluation. Furthermore, future research must address the classification of cognitive workload, including the measurement of excessive workload. In this respect, behavioural measures and embedded secondary tasks should be investigated. Finally, the literature is limited in its evaluation of the cognitive workload of the human–technology interfaces of systems, and more evidence is needed from studies that utilise realistic full-scale control room simulation, with the participation of professional control room operators.

### Acknowledgements

Thanks to the members of the of the Nuclear Energy Agency Halden Human Technology-Organisation Project's Program Review Group for their review and constructive comments.

### Disclosure statement

No potential conflict of interest was reported by the author(s).

### Funding

This work was funded by the Organisation for Economic Co-operation and Development (OECD) Nuclear Energy Agency, Halden Human Technology-Organisation (HTO) Project, Norway.

### ORCID

Per Øivind Braarud  <http://orcid.org/0000-0003-4545-4623>

### References

- Ackerman, P.L. 1987. "Individual Differences in Skill Learning: An Integration of Psychometric and Information Processing Perspectives." *Psychological Bulletin* 102 (1): 3–27. doi:10.1037/0033-2909.102.1.3.
- Al Harbi, M.A., A.R. Kim, I. Jang, P.H. Seong, S. Shirouzu, S. Katayama, and H.G. Kang. 2013. "Effects of Soft Control in the Nuclear Power Plants Emergency Operation Condition." *Annals of Nuclear Energy* 54: 184–191. doi:10.1016/j.anucene.2012.11.014.
- Arshad, S., Y. Wang, and F. Chen. 2013. "Analysing Mouse Activity for Cognitive Load Detection [Conference Session]." Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration, ACM, Adelaide, Australia, 115–118. doi:10.1145/2541016.2541083.
- Annett, J. 2002. "Subjective Rating Scales: Science or Art?" *Ergonomics* 45 (14): 966–987. doi:10.1080/00140130210166951.
- Battiste, V., and M.R. Bortolussi. 1988. "Transport Pilot Workload: A Comparison of Two Subjective Techniques." *Proceedings of the Human Factors Society Annual Meeting* 32 (2): 150–154. doi:10.1177/154193128803200232.
- Bittner, A. 1992. "Robust Testing and Evaluation of Systems: Framework, Approaches, and Illustrative Tools." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 34 (4): 477–484. doi:10.1177/001872089203400409.
- Boles, D.B., and L.P. Adair. 2001. "The Multiple Resources Questionnaire (MRQ)." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 45 (25): 1790–1794. doi:10.1177/154193120104502507.
- Borghini, G., L. Astolfi, G. Vecchiato, D. Mattia, and F. Babiloni. 2014. "Measuring Neurophysiological Signals in Aircraft Pilots and Car Drivers for the Assessment of Mental Workload, Fatigue and Drowsiness." *Neuroscience and Biobehavioral Reviews* 44: 58–75. doi:10.1016/j.neubiorev.2012.10.003.
- Bortolussi, M.R., R.B. Kantowitz, and S.G. Hart. 1986. "Measuring Pilot Workload in a Motion Base Trainer." *Applied Ergonomics* 17 (4): 278–283. doi:10.1016/0003-6870(86)90130-4.
- Bortolussi, M.R., S.G. Hart, and R.J. Shively. 1987. "Measuring Moment-to-Moment Pilot Workload Using Synchronous Presentations of Secondary Tasks in a Motion-Base Trainer." In *Proceedings of the Fourth Symposium on Aviation Psychology*. Columbus, OH: Ohio State University.
- Braarud, Per Øivind. 2021. "Investigating the Validity of Subjective Workload Rating (NASA TLX) AND Subjective Situation Awareness Rating (SART) FOR Cognitively Complex Human-machine Work." *International Journal of Industrial Ergonomics* 86: 103233. doi:10.1016/j.ergon.2021.103233.
- Braarud, P.Ø. 2000. *Subjective Task Complexity in the Control Room. Hwr-621*. Halden, Norway: OECD Halden Reactor Project.
- Braarud, P.Ø. 2020. "An Efficient Screening Technique for Acceptable Mental Workload Based on the NASA Task Load Index—Development and Application to Control Room Validation." *International Journal of Industrial Ergonomics* 76: 102904. doi:10.1016/j.ergon.2019.102904.
- Braarud, P.O., and B. Johansson. 2010. *Team Cognition in a Complex Accident Scenario. Hwr-955*. Halden, Norway: OECD Halden Reactor Project.
- Braarud, P.Ø., T. Bodal, J.E. Hulsund, M.N. Louka, C. Nihlwing, E. Nystad, H. Svengren, and E. Wingstedt. 2020. "An Investigation of Speech Features, Plant System Alarms, and Operator-System Interaction for the Classification of Operator Cognitive Workload during Dynamic Work." *Human Factors* 63 (5): 736–756. doi:10.1177/0018720820961730.
- Braarud, P.Ø, and G. Pignoni. 2022. *Human Factors Validation of Control Rooms - Insights and Experiences from Nuclear Industry Projects. Htor-020*. Halden, Norway: OECD Halden HTO Project.
- Braarud, P.Ø, and G. Pignoni. 2023. "Behavioral Indicators: An Approach for Assessing Control Room Operators' Excessive Cognitive Workload?" Proceedings of the AHFE 2023 International Conference on Applied Human Factors and Ergonomics. doi:10.54941/ahfe1003565.

- Brown, W.S., J.N. O'Hara, and J.C. Higgins. 2000. *Advanced Alarm Systems: Revision of Guidance and Its Technical Basis, NUREG/CR-6684*. Washington, DC: Nuclear Regulatory Commission.
- Cain, B. 2007. *A Review of the Mental Workload Literature. Technical Report RTO-TR-HFM-121-Part-II*. Toronto, Canada: Defence Research and Development.
- Carswell, C.M., C.H. Lio, R. Grant, M.I. Klein, D. Clarke, W.B. Seales, and S. Strup. 2010. "Hands-Free Administration of Subjective Workload Scales: Acceptability in a Surgical Training Environment." *Applied Ergonomics* 42 (1): 138–145. doi:10.1016/j.apergo.2010.06.003.
- Causse, M., Z.K. Chua, V. Peysakhovich, N. Del Campo, and N. Matton. 2017. "Mental Workload and Neural Efficiency Quantified in the Prefrontal Cortex Using fNIRS." *Scientific Reports* 7 (1): 5222. doi:10.1038/s41598-017-05378-x.
- Charfuelan, M., and G.J. Kruijff. 2013. "Classification of Speech under Stress and Cognitive Load in USAR Operations [Conference Session]." *Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference*, 508–512.
- Charles, R.L., and J. Nixon. 2019. "Measuring Mental Workload Using Physiological Measures: A Systematic Review." *Applied Ergonomics* 74: 221–232. doi:10.1016/j.apergo.2018.08.028.
- Chen, F., N. Ruiz, E. Choi, J. Epps, M.A. Khawaja, R. Taib, B. Yin, and Y. Wang. 2012. "Multimodal Behavior and Interaction as Indicators of Cognitive Load." *ACM Transactions on Interactive Intelligent Systems* 2 (4): 1–36. doi:10.1145/2395123.2395127.
- Chen, Y., S. Yan, and C.C. Tran. 2019. "Comprehensive Evaluation Method for User Interface Design in Nuclear Power Plant Based on Mental Workload." *Nuclear Engineering and Technology* 51 (2): 453–462. doi:10.1016/j.net.2018.10.010.
- Choi, M.K., S.M. Lee, J.S. Ha, and P.H. Seong. 2018. "Development of an EEG-Based Workload Measurement Method in Nuclear Power Plants." *Annals of Nuclear Energy* 111: 595–607. doi:10.1016/j.anucene.2017.08.032.
- Chuang, C., C.J. Lin, W. Shiang, T. Hsieh, and J. Liou. 2016. "Development of an Objective Mental Workload Assessment Tool Based on Rasmussen's Skill–Rule–Knowledge Framework." *Journal of Nuclear Science and Technology* 53 (1): 123–128. doi:10.1080/00223131.2015.1027156.
- Chung, Y., W. Yoon, and D. Min. 2009. "A Model-Based Framework for the Analysis of Team Communication in Nuclear Power Plants." *Reliability Engineering & System Safety* 94 (6): 1030–1040. doi:10.1016/j.res.2008.11.010.
- Colle, H.A., and G.B. Reid. 2005. "Estimating a Mental Workload Redline in a Simulated Air-to-Ground Combat Mission." *The International Journal of Aviation Psychology* 15 (4): 303–319. doi:10.1207/s15327108ijap1504\_1.
- Corwin, W.H., D.L. Sandry-Garza, M.H. Biferno, G.P. Boucek, A.L. Logan, J.E. Jonsson, and S.A. Metalis. 1989. *Assessment of Crew Workload Measurement Methods, Techniques, and Procedures: Process, Methods and Results. Report WRDC-TR-89-7006*. Wright-Patterson Air Force Base, OH: Wright Research and Development Centre, Air Force Systems Command.
- De Rivecourt, M., M.N. Kuperus, W.J. Post, and L.J. Mulder. 2008. "Cardiovascular and Eye Activity Measures as Indices for Momentary Changes in Mental Effort during Simulated Flight." *Ergonomics* 51 (9): 1295–1319. doi:10.1080/00140130802120267.
- De Waard, D., and B. Lewis Evans. 2014. "Self-Report Scales Alone Cannot Capture Mental Workload: A Reply to De Winter, Controversy in Human Factors Constructs and the Explosive Use of the NASA TLX: A Measurement Perspective." *Cognition, Technology & Work* 16 (3): 303–305. doi:10.1007/s10111-014-0277-z.
- Durantini, G., J.F. Gagnon, S. Tremblay, and F. Dehais. 2014. "Using near Infrared Spectroscopy and Heart Rate Variability to Detect Mental Overload." *Behavioural Brain Research* 259: 16–23. doi:10.1016/j.bbr.2013.10.042.
- Endsley, M.R., S.J. Selcon, T.D. Hardiman, and D.G. Croft. 1998. "A Comparative Evaluation of SAGAT and SART for Evaluations of Situation Awareness." *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*. Human Factors and Ergonomics Society, Santa Monica, CA, 82–86. doi:10.1177/154193129804200119.
- Farmer, E.W., and A. Brownson. 2003. *Review of Workload Measurement, Analysis and Interpretation Methods*. Brussels: European Organisation for the Safety of Air Navigation. (No. CARE-Integra-TRS-130-02-WP2).
- Gan, Y., X. Dong, Y. Zhang, X. Zhang, M. Jia, Z. Liu, and Z. Li. 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. doi:10.1016/j.ergon.2020.102974.
- Gao, Q., Y. Wang, F. Song, Z. Li, and X. Dong. 2013. "Mental Workload Measurement for Emergency Operating Procedures in Digital Nuclear Power Plants." *Ergonomics* 56 (7): 1070–1085. doi:10.1080/00140139.2013.790483.
- Gopher, D., and E. Donchin. 1986. "Workload - An Examination of the Concept." In *Handbook of Perception and Human Performance*, edited by K.R. Boff, L. Kaufman, and J.P. Thomas, Vol. II, Cognitive Processes and Performance. New York: Wiley & Sons.
- Grier, R.A. 2015. "How High Is High? A Meta-Analysis of NASA-TLX Global Workload Scores." *Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting*, 32, 150–154.
- Ha, C.H., J.H. Kim, S.J. Lee, and P.H. Seong. 2006. "Investigation on Relationship between Information Flow Rate and Mental Workload of Accident Diagnosis Tasks in NPPs." *IEEE Transactions on Nuclear Science* 53 (3): 1450–1459. doi:10.1109/TNS.2006.874189.
- Hancock, P.A., and J.S. Warm. 1989. "A Dynamic Model of Stress and Sustained Attention." *Human Factors*, 31 (5): 519–537. doi:10.1177/001872088903100503. 2625347
- Hancock, P.A. 2017. "Whither Workload? Mapping a Path for Its Future Development." In *Human Mental Workload: Models and Applications*, edited by L. Longo and C.M. Leva, 3–17. Berlin: Springer.
- Hancock, P.A., and G. Matthews. 2019. "Workload and Performance: Associations, Insensitivities, and Dissociations." *Human Factors* 61 (3): 374–392. doi:10.1177/0018720818809590.
- Hart, S. 2006. "Nasa-Task Load Index (Nasa-TLX); 20 Years Later." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50 (9): 904–908. doi:10.1177/154193120605000909.
- Hart, S.G., and L. Staveland. 1988. "Development of the NASA Task Load Index (TLX): Results of Empirical and Theoretical Research." In *Human Mental Workload*, edited by P. A. Hancock and N. Meshkati, 139–183. Amsterdam: North-Holland.
- Helmke, H., O. Ohneiser, T. Mühlhausen, and M. Wies. 2016. "Reducing Controller Workload with Automatic Speech Recognition." *IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, 1–10. doi:10.1109/DASC.2016.7778024.

- Hertzum, M. 2021. "Reference Values and Subscale Patterns for the Task Load Index (TLX): A Meta-Analytic Review." *Ergonomics* 64 (7): 869–878. doi:10.1080/00140139.2021.1876927.
- Hill, S.G., J.C. Byers, A.L. Zaklad, and R.E. Christ. 1989. *Subjective Workload Ratings of the LOS-F-H Mobile Air Defense Missile System in a Field Test Environment (Tech. Memo 5)*. Willow Grove, PA: Analytics, Inc.
- Hill, S.G., H.P. Iavecchia, J.C. Byers, A.C. Bittner, Jr., A.L. Zaklade, and R.E. Christ. 1992. "Comparison of Four Subjective Workload Rating Scales." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 34 (4): 429–439. doi:10.1177/001872089203400405.
- Hockey, G.R.J. 1997. "Compensatory Control in the Regulation of Human Performance under Stress and High Workload: A Cognitive-Energetical Framework." *Biological Psychology* 45: 73–79.
- Hsieh, M., S. Hwang, K. Liu, S. Liang, and C. Chuang. 2012. "A Decision Support System for Identifying Abnormal Operating Procedures in a Nuclear Power Plant." *Nuclear Engineering and Design* 249: 413–418. doi:10.1016/j.nucengdes.2012.04.009.
- Hsieh, M., M. Chiu, and S. Hwang. 2014. "An Interface Redesign for the Feed-Water System of the Advanced Boiling Water Reactor in a Nuclear Power Plant in Taiwan." *Journal of Nuclear Science and Technology* 51 (5): 720–729. doi:10.1080/00223131.2014.889584.
- Hsieh, M., M. Chiu, and S. Hwang. 2015. "An Optimal Range of Information Quantity on Computer-Based Procedure Interface Design in the Advanced Main Control Room." *Journal of Nuclear Science and Technology* 52 (5): 687–694. doi:10.1080/00223131.2014.977832.
- Huang, F.-H., S.-L. Hwang, T.-C. Yenn, Y.-C. Yu, C.-C. Hsu, and H.-W. Huang. 2006. "Evaluation and Comparison of Alarm Reset Modes in Advanced Control Room of Nuclear Power Plants." *Safety Science* 44 (10): 935–946. doi:10.1016/j.ssci.2006.06.002.
- Hwang, S.L., J. Lin, G. Liang, Y. Yau, T. Yenn, and C. Hsu. 2008. "Application Control Chart Concepts of Designing a Pre-Alarm System in the Nuclear Power Plant Control Room." *Nuclear Engineering and Design* 238 (12): 3522–3527. doi:10.1016/j.nucengdes.2008.07.011.
- Hwang, S.L., Y. Yau, Y. Lin, J.H. Chen, T. Huang, T. Yenn, and C. Hsu. 2008. "Predicting Work Performance in Nuclear Power Plants." *Safety Science* 46 (7): 1115–1124. doi:10.1016/j.ssci.2007.06.005.
- Hwang, S.L., S.F. Liang, T.Y.Y. Liu, Y.J. Yang, P.Y. Chen, and C.F. Chuang. 2009. "Evaluation of Human Factors in Interface Design in Main Control Room." *Nuclear Engineering and Design*. 239 (12): 3069–3075. doi:10.1016/j.nucengdes.2009.09.006.
- ISO. 2016. *ISO 6385. Ergonomics Principles in the Design of Work Systems*. Geneva: International Organization for Standardization.
- ISO. 2017. *EN ISO 10075-1:2017. Ergonomic Principles Related to Mental workload - Part 1: General Issues and Concepts, Terms and Definitions*. Brussels: CEN
- Jimmieson, N.L., and D.J. Terry. 1999. "The Moderating Role of Task Characteristics in Determining Responses to a Stressful Work Simulation." *Journal of Organizational Behavior* 20 (5): 709–736. doi:10.1002/(SICI)1099-1379(199909)20:5<709::AID-JOB954>3.0.CO;2-7.
- Jordan, C.S. 1992. *Experimental Study of the Effect of an Instantaneous Self-Assessment Workload Recorder on Task Performance. DRA Technical Memorandum (CAD5) 92011*. Portsmouth: DRA Maritime Command Control Division.
- Jou, Y., T. Yenn, C.J. Lin, C. Yang, and S. Lin. 2009. "Evaluation of Mental Workload in Automation Design for a Main Control Room Task." International Conference on Networking, Sensing and Control, 313–317. doi:10.1109/ICNSC.2009.4919293.
- Khawaja, M.A., F. Chen, and N. Marcus. 2012. "Analysis of Collaborative Communication for Linguistic Cues of Cognitive Load." *Human Factors* 54 (4): 518–529. doi:10.1177/0018720811431258.
- Kim, S., Y. Kim, and W. Jung. 2014. "Operator's Cognitive, Communicative and Operative Activities Based Workload Measurement of Advanced Main Control Room." *Annals of Nuclear Energy*. 72: 120–129. doi:10.1016/j.anucene.2014.04.019.
- Lau, N., G.A. Jamieson, G. Skraaning, and C.M. Burns. 2008. "Ecological Interface Design in the Nuclear Domain: An Empirical Evaluation of Ecological Displays for the Secondary Subsystems of a Boiling Water Reactor Plant Simulator." *IEEE Transactions on Nuclear Science* 55 (6): 3597–3610. doi:10.1109/TNS.2008.2005725.
- Lin, C.J., T. Yenn, and C. Yang. 2010. "Automation Design in Advanced Control Rooms of the Modernized Nuclear Power Plants." *Safety Science* 48 (1): 63–71. doi:10.1016/j.ssci.2009.05.005.
- Lin, C.J., T.-L. Hsieh, and S.-F. Lin. 2013. "Development of Staffing Evaluation Principle for Advanced Main Control Room and the Effect on Situation Awareness and Mental Workload." *Nuclear Engineering and Design* 265: 137–144. doi:10.1016/j.nucengdes.2013.08.056.
- Lin, J.T., Y.C. Chen, S.C. Wu, and S.L. Hwang. 2017. "Development and Evaluation of a Monitoring-Aid System for a Nuclear Power Plant in Control Room System Manipulation." *Work* 57 (4): 611–625. doi:10.3233/WOR-172587.
- Longo, L., C.D. Wickens, G. Hancock, and P.A. Hancock. 2022. "Human Mental Workload: A Survey and a Novel Inclusive Definition." *Frontiers in Psychology* 13: 883321. doi:10.3389/fpsyg.2022.883321.
- Lysaght, R.J., S.G. Hill, A.O. Dick, B.D. Plamondon, P.M. Linton, W.W. Wierwille, A.L. Zaklad, A.C. Bittner, Jr., and R.J. Wherry, Jr. 1989. *Operator Workload: Comprehensive Review and Evaluation of Operator Workload Methodologies (ARI Technical Report 851)*. Fort Bliss, TX: United States Army Research Institute.
- McCracken, J.H. and T.B. Aldrich. 1984. "Analyses of Selected LHX Mission Functions: Implications for Operator Workload and System Automation Goals." Technical Note AD-A232-330.
- McDonald, A.D., T.K. Ferris, and T.A. Wiener. 2020. "Classification of Driver Distraction: A Comprehensive Analysis of Feature Generation, Machine Learning, and Input Measures." *Human Factors* 62 (6): 1019–1035. doi:10.1177/0018720819856454.
- McKendrick, R., B. Feest, A. Harwood, and B. Falcone. 2019. "Theories and Methods for Labeling Cognitive Workload: Classification and Transfer Learning." *Frontiers in Human Neuroscience* 13: 295. doi:10.3389/fnhum.2019.00295.
- Megaw, T. 2005. "The Definition and Measurement of Mental Workload." In *Evaluation of Human Work*, edited by J.R. Wilson and N. Corlett, 3rd ed., 525–551. Abingdon-on-Thames: Taylor & Francis Group.



- Messick, S. 1990. *Validity of Test Interpretation and Use. Research Report 90-11*. Princeton, NJ: Education Testing Service.
- Miyake, S., S. Yamada, T. Shoji, Y. Takae, N. Kuge, and T. Yamamura. 2009. "Physiological Responses to Workload Change. A Test/Retest Examination." *Applied Ergonomics* 40 (6): 987–996. doi:10.1016/j.apergo.2009.02.005.
- Moher, D., A. Liberati, J. Tetzlaff, and D.G. Altman. 2009. "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement." *Annals of Internal Medicine* 151 (4): 264–269, W64. doi:10.7326/0003-4819-151-4-200908180-00135.
- Moray, N. 1988. "Mental Workload since 1979." *International Review of Ergonomics* 2: 123–150.
- Muckler, F.A., and S.A. Seven. 1992. "Selecting Performance Measures: 'Objective' versus 'Subjective.'" *Human Factors: The Journal of the Human Factors and Ergonomics Society* 34 (4): 441–455. doi:10.1177/001872089203400406.
- Mulder, G. 1979. Mental Workload. NATO Conference Series." In *Mental Load, Mental Effort and Attention*, edited by N. Moray. Boston, MA: Springer.
- Mulder, G., and L.J. Mulder. 1981. "Information Processing and Cardiovascular Control." *Psychophysiology* 18 (4): 392–402. doi:10.1111/j.1469-8986.1981.tb02470.x.
- Mulder, G., L.J.M. Mulder, T.F. Meijman, J.B.P. Veldman, and A.M.v. Roon. 2000. "A Psychophysiological Approach to Working Conditions." In *Engineering Psychophysiology: Issues and Applications*, edited by R.W. Backs and W. Boucsein, 139–159. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Neerincx, M.A. 2003. "Cognitive Task Load Analysis: Allocating Tasks and Designing Support." In *Handbook of Cognitive Task Design*, edited by E. Hollnagel, 283–305. Mahwah, NJ: Lawrence Erlbaum Associates Publishers. doi:10.1201/9781410607775.
- O'Donnell, R.D., and F.T. Eggemeier. 1986. "Workload Assessment Methodology." In *Handbook of Perception and Human Performance*, edited by K.R. Boff, L. Kaufman, J.P. Thomas, Vol. II, 1–49. New York: Wiley & Sons.
- O'Hara, J., J. Higgins, S. Fleger, and P. Pieringer. 2012. *Human Factors Engineering Program Review Model (NUREG-0711, Rev. 3)*. Washington, DC: U.S. Nuclear Regulatory Commission.
- OECD NEA. 2017. "Human Factors Validation of Nuclear Power Plant Control Room Designs and Modifications." Proceedings of the Expert Workshop Charlotte, United States 19-21 February 2015. NEA/CSNI/R(2016)17, The Nuclear Energy Agency (NEA), Paris, France.
- Or, C.K., and V.G. Duffy. 2007. "Development of a Facial Skin Temperature-Based Methodology for Non-Intrusive Mental Workload Measurement." *Occupational Ergonomics* 7 (2): 83–94. doi:10.3233/OER-2007-7202.
- Parasuraman, R. 2003. "Neuroergonomics: Research and Practice." *Theoretical Issues in Ergonomics Science* 4 (1-2): 5–20. doi:10.1080/14639220210199753.
- Parasuraman, R., T.B. Sheridan, and C.D. Wickens. 2008. "Situation Awareness, Mental Workload, and Trust in Automation: viable, Empirically Supported Cognitive Engineering Constructs." *Journal of Cognitive Engineering and Decision Making* 2 (2): 140–160. doi:10.1518/155534308X284417.
- Park, J., D. Lee, W. Jung, and J. Kim. 2017. "An Experimental Investigation on Relationship between PSFs and Operator Performances in the Digital Main Control Room." *Annals of Nuclear Energy* 101: 58–68. doi:10.1016/j.anucene.2016.10.020.
- Park, J., W. Jung, and J. Kim. 2020. "Inter-Relationships between Performance Shaping Factors for Human Reliability Analysis of Nuclear Power Plants." *Nuclear Engineering and Technology* 52 (1): 87–100. doi:10.1016/j.net.2019.07.004.
- Perrow, C. 1984. *Normal Accidents: Living with High-Risk Technologies*. New York: Basic Books.
- Perry, C.M., N. Segall, and D.B. Kaber. 2005. "Measurement of Situation Awareness Effects of Adaptive Automation of Air Traffic Control Information Processing Functions." 2005 International Symposium on Aviation Psychology, 566–571.
- Pickup, L., J. Wilson, and E. Lowe. 2010. "The Operational Demand Evaluation Checklist (ODEC) of Workload for Railway Signaling." *Applied Ergonomics* 41 (3): 393–402. doi:10.1016/j.apergo.2009.09.003.
- Pimenta, A., D. Carneiro, J. Neves, and P. Novais. 2016. "A Neural Network to Classify Fatigue from Human–Computer Interaction." *Neurocomputing* 172: 413–426. doi:10.1016/j.neucom.2015.03.105.
- Rasmussen, J. 1986. *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering*. New York, NY: Elsevier.
- Reid, G.B., and T.E. Nygren. 1988. "The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload." *Advances in Psychology* 52: 185–218.
- Reinerman-Jones, L., S. Guznov, J. Tyson, A. D'Agostino, and N. Hughes. 2015. *Workload, Situation Awareness, and Teamwork (NUREG/CR-7190)*. Washington, DC: U.S. Nuclear Regulatory Commission.
- Reinerman-Jones, L., G. Matthews, and J.E. Mercado. 2016. "Detection Tasks in Nuclear Power Plant Operation: Vigilance Decrement and Physiological Workload Monitoring." *Safety Science* 88: 97–107. doi:10.1016/j.ssci.2016.05.002.
- Reinerman, L., J. Mercado, J.L. Szalma, and P.A. Hancock. 2020. "Understanding Individualistic Response Patterns When Assessing Expert Operators on Nuclear Power Plant Control Tasks." *Ergonomics* 63 (4): 440–460. doi:10.1080/00140139.2019.1677946.
- Rubio, Susana, Eva Díaz, Jesús Martín, and José M. Puente. 2004. "Evaluation of Subjective Mental Workload: A Comparison of Swat, Nasa-Tlx, and Workload Profile Methods." *Applied Psychology* 53 (1): 61–86. doi:10.1111/j.1464-0597.2004.00161.x.
- Shiffrin, R.M., and W. Schneider. 1977. "Controlled and Automatic Human Information Processing: II. Perceptual Learning, Automatic Attending and a General Theory." *Psychological Review* 84 (2): 127–190. doi:10.1037/0033-295X.84.2.127.
- Simonsen, E., and A.L. Osvalder. 2018. "Categories of Measures to Guide Choice of Human Factors Methods for Nuclear Power Plant Control Room Evaluation." *Safety Science* 102: 101–109. doi:10.1016/j.ssci.2017.10.006.
- Slocum, G.K., B.H. Williges, and S.N. Roscoe. 1971. "Meaningful Shape Coding for Aircraft Switch Knobs." *Aviation Research Monographs* 1 (3): 27–40.
- Solovey, E.T., M. Zec, E.A.G. Perez, B. Reimer, and B. Mehler. 2014. "Classifying Driver Workload Using Physiological and Driving Performance Data: Two Field Studies [Conference Session]." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, NY, USA, ACM, 4057–4066.
- Tobaruela, G., W. Schuster, A. Majumdar, W.Y. Ochieng, L. Martinez, and P. Hendrickx. 2014. "A Method to Estimate Air Traffic Controller Mental Workload Based on Traffic Clearances." *Journal of Air Transport Management* 39: 59–71. doi:10.1016/j.jairtraman.2014.04.002.

- Tsang, P.S., and M.A. Vidulich. 2006. "Mental Workload and Situation Awareness." In *Handbook of Human Factors and Ergonomics*, edited by G. Salvendy, 243–268. New York: John Wiley & Sons, Inc. doi:10.1002/0470048204.ch9.
- Veltman, J.A., and A.W.K. Gaillard. 1998. "Physiological Workload Reactions to Increasing Levels of Task Difficulty." *Ergonomics* 41 (5): 656–669. doi:10.1080/001401398186829.
- Vicente, K.J. 1999. *Cognitive Work Analysis: Towards Safe, Productive, and Healthy Computer-Based Work*. Mahwah, NJ: LEA.
- Vidulich, M.A., and P.S. Tsang. 1986. "Techniques of Subjective Workload Assessment: A Comparison of SWAT AND THE NASA-Bipolar Methods." *Ergonomics* 29 (11): 1385–1398. doi:10.1080/00140138608967253.
- Vidulich, M.A. 2000. "The Relationship between Mental Workload and Situation Awareness." Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 460–463.
- Vidulich, M.A. and P.S. Tsang. 2015. "The Confluence of Situation Awareness and Mental Workload for Adaptable Human-Machine Systems." *Journal of Cognitive Engineering and Decision Making* 9 (1): 95–97. doi:10.1177/1555343414554805.
- Wickens, C.D. 2000. "Workload and Situation Awareness." In *Stress, Workload, and Fatigue*, edited by P.A. Hancock and P.A. Desmond, 443–450. Mahwah, NJ: Lawrence Erlbaum Associates.
- Wickens, C.D., J.G. Hollands, S. Banbury, and R. Parasuraman. 2012. *Engineering Psychology and Human Performance*. 4th ed. London: Psychology Press.
- Wierwille, W.W., and J.G. Casali. 1983. "A Validated Rating Scale for Global Mental Workload Measurement Application." *Proceedings of the Human Factors Society Annual Meeting* 27 (2): 129–133. Santa Monica, CA: Human Factors Society. doi:10.1177/154193128302700203.
- Wilson, G.F., and C.A. Russell. 2003. "Real-Time Assessment of Mental Workload Using Psychophysiological Measures and Artificial Neural Networks." *Human Factors* 45 (4): 635–643. doi:10.1518/hfes.45.4.635.27088.
- Woods, D. 1988. "Coping with Complexity: The Psychology of Human Behaviour in Complex Systems."
- World Nuclear Association. 2022. <https://world-nuclear.org/information-library/current-and-future-generation/world-energy-needs-and-nuclear-power.aspx>
- Wu, X., and Z. Li. 2013. "Secondary Task Method for Workload Measurement in Alarm Monitoring and Identification Tasks [Conference Session]." International Conference on Cross-Cultural Design, Las Vegas, NV, Springer Berlin Heidelberg, 346–354.
- Wu, X., X. Yuan, Z. Li, F. Song, and W. Sang. 2016. "Validation of "Alarm Bar" Alternative Interface for Digital Control Panel Design: A Preliminary Experimental Study." *International Journal of Industrial Ergonomics* 51: 43–51. doi:10.1016/j.ergon.2014.11.002.
- Wu, Y., Z. Liu, M. Jia, C.C. Tran, and S. Yan. 2020. "Using Artificial Neural Networks for Predicting Mental Workload in Nuclear Power Plants Based on Eye Tracking." *Nuclear Technology* 206 (1): 94–106. doi:10.1080/00295450.2019.1620055.
- Xie, B., and G. Salvendy. 2000. "Prediction of Mental Workload in Single and Multiple Tasks Environments." *International Journal of Cognitive Ergonomics* 4 (3): 213–242. doi:10.1207/S15327566IJCE0403\_3.
- Xu, S., F. Song, Z. Li, Q. Zhao, W. Luo, X. He, and G. Salvendy. 2008. "An Ergonomics Study of Computerized Emergency Operating Procedures: Presentation Style, Task Complexity, and Training Level." *Reliability Engineering & System Safety*. 93 (10): 1500–1511. doi:10.1016/j.res.2007.09.006.
- Yan, S., C.C. Tran, Y. Chen, K. Tan, and J.L. Habiaryemye. 2017. "Effect of User Interface Layout on the Operators' Mental Workload in Emergency Operating Procedures in Nuclear Power Plants." *Nuclear Engineering and Design* 322: 266–276. doi:10.1016/j.nucengdes.2017.07.012.
- Yang, C.W., L.C. Yang, T.C. Cheng, Y.T. Jou, and S.W. Chiou. 2012. "Assessing Mental Workload and Situation Awareness in the Evaluation of Computerized Procedures in the Main Control Room." *Nuclear Engineering and Design* 250: 713–719. doi:10.1016/j.nucengdes.2012.05.038.
- Yeh, Y.Y., and C.D. Wickens. 1988. "Dissociation of Performance and Subjective Measures of Workload." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 30 (1): 111–120. doi:10.1177/001872088803000110.
- Young, M.S., K.A. Brookhuis, C.D. Wickens, and P.A. Hancock. 2015. "State of Science: Mental Workload in Ergonomics." *Ergonomics* 58 (1): 1–17. doi:10.1080/00140139.2014.956151.
- Zijlstra, F.R.H. 1993. "Efficiency in Work Behavior. A Design Approach for Modern Tools." Ph.D. thesis, Delft University of Technology, Delft University Press, Delft.