Human Performance Measures for the Evaluation of Nuclear Power Plant Control Room Interfaces: a Systematic Review

Jie Xu[1], Shilo Anders[1], Arisa Pruttianan[2], Daniel J. France[1], Nathan Lau[2], Julie A. Adams[3], and Matthew B. Weinger[1]

[1] Center for Research and Innovation in System Safety, Vanderbilt University Medical Center
1211 21st Avenue S., MAB Suite 732, Nashville, TN 37212
jie.xu@vanderbilt.edu; shilo.anders@vanderbilt.edu; dan.france@vanderbilt.edu; matt.weinger@vanderbilt.edu

[2] Department of Industrial & Systems Engineering, Virginia Tech
250 Durham Hall, 1145 Perry Street, Blacksburg, VA 24061
apruttia@vt.edu; nathan.lau@vt.edu

[3] School of Electrical Engineering and Computer Science at Oregon State University
1148 Kelley Engineering Center, Corvallis, OR 97331
julie.a.adams@oregonstate.edu

ABSTRACT

The purpose of this study was to review the literature related to measuring human performance in the design and evaluation of the human-system interface (HSI) of nuclear power plant (NPP) control rooms and synthesize lessons learned and future directions for research and practice in this field. The design and evaluation of digital or hybrid HSI requires comprehensive measurement of human performance to ensure quality and safety of the nuclear power process control. Human factors evaluation frameworks have been proposed to guide the measurement of human performance constructs. However, there are few guides for selecting appropriate measures for those constructs. A systematic review of empirical research literature was conducted. Studies were identified through academic publication databases and grey literature searches. Eligible studies were coded and information regarding human performance measures was extracted and analyzed. A total of 80 studies met the criteria. Human performance measures extracted from the reviewed studies included measures of: behavioral performance; workload; situation awareness; human-human or human-machine cooperation; other cognitive performance indicators; and system performance. Behavioral performance (speed and accuracy), workload, and situation awareness were the most frequently measured constructs with validated and domain-specific instruments. The results of the literature review provided an overview of the human performance constructs and their corresponding measures for NPP control room HSI evaluation. These results can be used as an initial guide to establish appropriate and comprehensive measurements of human performance for such evaluations. Future directions in measure development are also discussed.

Key Words: Human factors, human performance measures, human-system interface, systematic review, control room design

1 INTRODUCTION

The United States is the largest producer of nuclear power in the world and over 19% of the country's total electrical output is produced by nuclear reactors [1]. As most of the nuclear power plants (NPPs) in the US are reaching their 40-year operating license and are applying for license extensions, it is critical to ensure these NPPs continue to operate efficiently and safely in the future [2]. Control room modernizations, which include fully or partially digital human-system interface (HSI) upgrades, are performed to replace the aging analog systems in many plants [3; 4]. A survey of nuclear utility representatives indicated that the majority of them prefer a modernized HSI control room [5]. The motivation for HSI digital upgrades may include: the high cost of maintaining the existing analog systems; the reduced need for operator manual tasks with the introduction of automation; the drive towards industry best practices; and the influence of regulatory recommendations [6].

However, the introduction of new HSI technology, such as computer-based procedures, automated control, and handheld digital devices, may have unintended consequences on human performance. Human performance issues related to HSI design deficiencies were cited as causes of NPP safety-related incidents. For example, an analysis of the Chernobyl incident indicated that operators had difficulties in understanding the full effects of their actions due to system complexity [7]. In the Three Mile Island incident, operator performance was hindered as key information was hidden from the operators by other information or by physical objects; Some of the key controls were difficult to reach due to their physical layout [7; 8]. Thus, it is necessary to integrate human factors processes into the entire HSI design and evaluation process, from planning and analysis, design, verification and validation (V&V), to implementation and operation [9; 10].

Human factors assessment guidelines and frameworks have been proposed by researchers [11-13] and regulators [9]. Yet, guidance on the use of specific human performance measures is still not readily available [14]. In fact, a recent report prepared by the Idaho National Laboratory stated that there is a growing need for guidance on the development of new and application of existing human performance measures for design and evaluation of modern NPP HSI [10]. The purpose of this systematic review was to take the first step in developing such guidance by reviewing human performance measures used in NPP control room HSI design and evaluation and synthesizing lessons learned and future directions.

In this review, human performance is broadly defined as the ability of an operator or a crew to accomplish task requirements [7]. Evaluation of this "ability" can be performed on different dimensions. According to multiple human performance evaluation frameworks [9; 11-13], at least three dimensions of human performance are particularly relevant to human-system interaction: behavioral/task/direct performance, workload, and situation awareness (SA). Additional dimensions may include plant/system performance [9; 12; 13] and teamwork [12; 13]. This review included measures of all dimensions of human performance.

The usefulness of a human performance measure is limited by the extent to which it measures what it is designed to measure (*validity*), the degree to which it produces stable and consistent results (*reliability*), and the ability to which it provides appropriate level of detail for meaningful analysis (*sensitivity*). Research in psychometrics has established widely accepted criteria for evaluating whether a measure has achieved acceptable levels of validity, reliability, and sensitivity [15-17]. Specifically, different facets of validity and reliability have been defined. For example, facets of reliability include internal consistency, inter-rater reliability, test-retest reliability, and parallel forms reliability [18]. Criteria exist for the evaluation of each of the facets. For example, a Cronbach's alpha above 0.70 indicates acceptable internal consistency [15; 17]. Human performance measures developed for use in the NPP domain should be evaluated according to the established criteria for validity, reliability, and sensitivity [19]. Thus, this review aimed to collect evidence from prior NPP HSI human performance studies and to assess the psychometric properties of relevant human performance measures.

2 METHODS

This review included studies that: (1) Were experimental or observational studies involving human participants; (2) Evaluated one or more components of HSI (*HSI Evaluation Studies*) or developed methods or measures for such evaluation (*Methodology Studies*); (3) Were conducted in a real or simulated NPP control room; (4) Measured human performance and reported the resultant data. Articles were reviewed to extract the relevant studies. An article was included if it described one or more eligible studies, was written in English, and had full text available to the research team.

The terms used for the literature search included human performance related terms (e.g., “operator performance”, “crew performance”, “workload”, and “situation awareness”) and process control domain related terms (e.g., “nuclear power” and “power plant”). IEEExplore, Ei Compendex, PubMed, and International Nuclear Information System (INIS) repository were searched to identify eligible articles. A Human factors and process control expert also provided a list of potentially relevant articles. All the articles were assessed for eligibility by title and abstract screening. Addition articles were identified through reference list search and Google Scholar “cite by” search based on the eligible articles. Full texts of the included articles were reviewed to identify eligible studies. Eligible studies were used as the unit of analysis in the review. **Figure 1** depicts the flow diagram of the literature search and review.

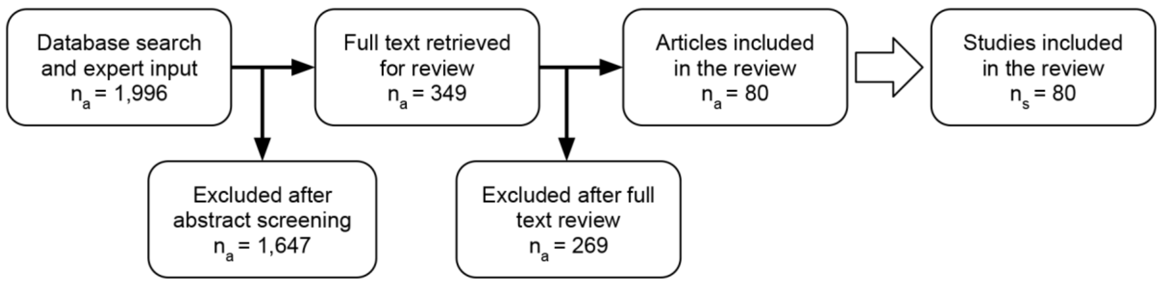

Figure 1. Identification and selection of studies for review. n_a, Number of articles. n_s, number of studies.

Data from all the included articles and studies were extracted using web-based standardized data extraction forms, which included fields about article reference information (e.g., title, authors, year of publication), study attributes (e.g., experimental vs. observation, study population, sample size), setting (e.g., real world vs. simulation, HSI components evaluated), variables and measures, and main results. Detailed information, such as data collection method, administration time, and psychometric property for each human performance measures were extracted from each study. . Psychometric properties of the measures were then evaluated against widely accepted criteria [15-17] to assess if a measure achieved acceptable levels. Properties measured included internal consistency, inter-rater reliability, content validity, structural validity, convergent validity, concurrent validity, and sensitivity.

3 RESULTS

3.1 Search Results and Study Characteristics

A total of 1,996 articles were retrieved from the initial literature search, expert input, reference list search, and Google Scholar “cite by” search after removal of duplicates (**Figure 1**). After screening of title and abstract, 1,647 articles were excluded. Full texts of 349 articles were retrieved for review. Based on the inclusion criteria, 269 articles were excluded and 80 were included in this review. Among the articles included, some described more than one study (up to four studies). In some cases, multiple articles described the same study, for example, a technical report, a conference proceeding, and a journal article may describe the same study in different levels of detail. A total of 80 studies were extracted from the 80 articles. All 80 studies were included in this review.

Among the included studies, there were 57 HSI Evaluation Studies and 25 Methodology Studies (two studies were categorized as both as they evaluated HSI and developed new measures). The majority of the studies were experimental studies (69, 86%). Most studies were conducted in a simulated environment (78, 98%). Study participants were usually operators (46, 58%) or university students (31, 39%). A quarter the studies used a small sample size of 10 or less individuals (20, 25%) and the majority used 20 or less (58, 73%). HSI components involved in the studies included: overview display; computer-based procedure; alarm; control interface; decision aid; and handheld device. See **Table I** for more details of the characteristics of the included studies.

Table I. Characteristics of the studies included in this review.

	HSI Evaluation Studies	Methodology Studies	All studies
Total Number	57	25	80[1]
Study Design			
Experimental	51	20	69[1]
Observational	5	5	10
Quasi-experimental	1	0	1
Study Setting			
Simulator	56	24	78[1]
Real world	1	1	2
Sample Population			
Operator	31	17	46[1]
Students	24	7	31
Other (e.g., research staff)	2	1	3
Sample Size			
10 or less	16	4	20
11-20	24	16	38[1]
20 or more	17	4	21
HSI component(s) involved			
Overview display	13	3	16
Computer-based procedure	11	2	13
Alarm	6	1	7
Control interface	5	0	5
Decision aid	3	1	4
Handheld device	1	0	1
Multiple components	18	18	34[1]

[1] The number in "all studies" does not equal to the sum of "HSI Evaluation Studies" and "Methodology Studies" since two studies were categorized as both kinds of studies.

3.2 Human Performance Measures

The most frequently used human performance measures in the reviewed studies included behavioral performance, workload, and SA. See **Table II** for a list of all the measures reviewed in these categories and their corresponding human performance constructs, measure types, usages, and references.

Table II. Behavioral performance, workload, and situation awareness measures reviewed in this study.

Construct	Measure Type	Measure	Usage[1]	References
Behavioral Performance				
Operation efficiency	*Behavioral indicator*	Completion time	26 (23, 3)	[20-41]
		Response time	10 (8, 2)	[30; 38; 42-48]
		Action frequency	2 (2, 0)	[29; 43]
		Completion rate	1 (0, 1)	[49]
		Mean Navigation Level (MNL)	1 (1, 0)	[35]
Operation accuracy	*Behavioral indicator*	Error rate	22 (19, 3)	[21; 23; 26; 28; 31-35; 38-41; 43; 50-55]
		Completion rate	6 (6, 0)	[20; 26; 36; 52; 56; 57]
Overall performance	*Observer rating*	Operator Performance Assessment System (OPAS)	10 (4, 6)	[44; 58-67]
		Other expert-rated performance (various instruments)	10 (6, 4)	[20; 51; 61; 68-73]
	Self-report rating	Performance Evaluation Matrix (PEM) questionnaire	1 (0, 1)	[74]
		Other self-reported performance (various instruments)	9 (4, 5)	[22; 42; 60-62; 66; 71; 74]
Workload				
Workload – individual	*Self-report rating*	NASA Task Load Index (TLX)	35 (25, 10)	[21; 23-25; 30-35; 38; 39; 43; 48; 49; 51; 54; 55; 57; 67; 68; 73; 75-86]
		Modified Cooper Harper (MCH) scale	1 (1, 0)	[77]
		Other self-reported workload (various instruments)	3 (2, 1)	[22; 41; 69]
	Secondary task performance	Error rate	6 (5, 1)	[24; 30; 56; 75; 80; 83]
		Response time	3 (3, 0)	[33; 56; 80]
	Physiological indicator	Cardiac activity – heart rate	3 (2, 1)	[22; 33; 82]
		Cardiac activity – heart rate variability	3 (1, 2)	[39; 81; 82]
		Eye tracking – blink rate	3 (2, 1)	[39; 77; 87]
		Eye tracking – blink duration	1 (0, 1)	[39]
		Eye tracking – number of fixations	1 (1, 0)	[77]
		Eye tracking – duration of fixations	3 (1, 2)	[77; 85]
		Eye tracking – pupil size	2 (1, 1)	[39; 87]
		Electroencephalogram (EEG)	1 (0, 1)	[82]
		Integrated score of physiological signals	2 (0, 2)	[39; 83]
	Behavioral indicator	COgnitive, Communicative, and Operational Activities (COCOA)	2 (1, 1)	[88; 89]
		Activity level	3 (0, 3)	[85; 86]
		GOMS-KLM time estimation and workload index	1 (0, 1)	[39]
	Qualitative method	Observation and interview	2 (2, 0)	[90; 91]
Workload – team	*Self-report rating*	Team workload assessment (TWA)	1 (0, 1)	[49]
Situation Awareness (SA)				
SA - individual	*Freeze probe technique*	Situation Awareness Control Room Inventory (SACRI)	11 (3, 8)	[36; 47; 68; 73; 85; 92; 93]
		Situation Awareness Global Assessment Techniques (SAGAT)	2 (2, 0)	[78; 80]
		Process Overview	5 (2, 3)	[62; 64; 65; 94-96]
		Halden Open Probe Elicitation (HOPE)	3 (2, 1)	[64; 65; 94; 95]
	Self-report rating	Situational Awareness Rating Technique (SART)	6 (6, 0)	[36; 43; 44; 56; 75; 79]
		Important Parameter Assessment Questionnaire (IPAQ)	1 (1, 0)	[47]
		Other self-reported situation awareness (various instruments)	4 (4, 0)	[25; 52; 69]
	Observer rating	Automation and Scenario Understanding Rating	1 (0, 1)	[64; 65]

Construct	Measure Type	Measure	Usage[1]	References
		Scales (ASURS)		
	Physiological indicator	Visual Indicator of Situation Awareness (VISA)	2 (0, 2)	[85]
	Qualitative method	Observation and interview	1 (1, 0)	[90]
SA - team	*Self-report rating*	Team situation awareness questionnaire	1 (1, 0)	[30]

[1] The numbers indicate: total number of times used (number of times used in HSI Evaluation Studies, number of times used in Methodology Studies).

The most widely used behavioral indicators of behavioral performance included completion time, response time, error rate, and completion rate. However, their definitions may vary in different studies. For example, completion time may be defined as time to complete a scenario [25], a procedure [38], or key actions [23]. Many observer-rated and self-reported measures were used, however, most of them were either not validated or not described in sufficient detail. The notable exception was the Operator Performance Assessment System (OPAS), which was a well-validated domain-specific, task-analysis driven, process-expert-rated measure [61].

NASA Task Load Index (NASA TLX) was the most widely used workload measure. The use rates of all the other workload measures were relatively low. One notable finding is that many physiological workload indicators were used but there appears to be no consensus on which one(s) to use.

Situation Awareness Global Assessment Techniques (SAGAT) and Situational Awareness Rating Technique (SART) are widely-used SA measures across different domains [97] and they were also used in the reviewed studies. However, the domain-specific measures were more popular, such as Situation Awareness Control Room Inventory (SACRI), Process Overview, and Halden Open Probe Elicitation (HOPE).

A number of measures of other dimensions of human performance also emerged in the review. These measures were categorized as human-human or human-machine cooperation, other cognitive indicators, and system performances (see **Table III**). Most of these measures were not frequently used, with the notable exception of the Subjective Task Complexity Scale, which was sometimes used as a measure of workload [58; 59].

Table III. Human-human or human-machine cooperation, other cognitive performance, and system performance measures reviewed in this study.

Construct	Measure Type	Measure	Usage[1]	References
Human-Human or Human-Machine Cooperation				
Teamwork	*Observer rating*	Behaviorally Anchored Rating Scale (BARS)	3 (1, 2)	[68; 73]
		Crew teamwork performance rating	1 (1, 0)	[71]
	Self-report rating	Teamwork skill rating questionnaire	2 (2, 0)	[26]
Communication – effectiveness	*Qualitative method*	Observation and interview	1 (1, 0)	[90]
Communication – frequency	*Observer rating*	Command frequency	1 (1, 0)	[56]
Human-human cooperation	*Self-report rating*	Halden Cooperation Scale (HCS) - human-human	3 (1, 2)	[44; 66]
Human-machine cooperation	*Self-report rating*	Halden Cooperation Scale (HCS) - human-automation	3 (1, 2)	[44; 66]
Trust in automation	*Self-report rating*	Trust in automation scale - general	2 (0, 2)	[66; 71]
		Trust in automation scale - specific	3 (1, 2)	[44; 66; 71]
		Operators' Trust in the Automation (OTA)	1 (1, 0)	[42]
Out-of-the-loop	*Self-report rating*	Out-of-the-loop (OOTP) questionnaire	1 (1, 0)	[47]

Construct	Measure Type	Measure	Usage[1]	References
Other Cognitive Performance Indicators				
Attention	*Physiological indicator*	Fixation to Importance Ratio (FIR)	3 (0, 3)	[98; 99]
		Selective Attention Effectiveness (SAE)	2 (0, 2)	[98; 99]
Perception	*Self-report rating*	Perception effectiveness scale	1 (1, 0)	[76]
Cognition – general	*Self-report rating*	Cognition effectiveness scale	1 (1, 0)	[76]
Cognition – information search	*Physiological indicator*	Eye tracking - number of fixations	1 (1, 0)	[87]
Cognition – comprehension	*Physiological indicator*	Eye tracking - duration of fixations	1 (1, 0)	[87]
	Self-report rating	Comprehension test	1 (1, 0)	[57]
Cognition – problem-solving	*Qualitative method*	Performance flow diagram	1 (1, 0)	[20]
Cognition – mental model	*Self-report rating*	Mental model accuracy score	1 (0, 1)	[98]
Perceived complexity	*Self-report rating*	Subject Task Complexity Scale	10 (4, 6)	[44; 58-60; 62; 73; 86]
Metacognition	*Observation and self-report rating*	Metacognitive bias	1 (1, 0)	[60]
		Metacognitive accuracy	1 (1, 0)	[60]
Responses – manual	*Self-report rating*	Manual responses effectiveness	1 (1, 0)	[76]
Responses – vocal	*Self-report rating*	Vocal responses effectiveness	1 (1, 0)	[76]
System Performance				
Plant Performance	*Observer rating*	Activity PROfiling MatriX (APROX)	1 (0, 1)	[48]
	Plant parameters	Plant Performance Assessment System (PPAS)	2 (0, 2)	[73; 100]
		Cool-down and stabilization	1 (1, 0)	[68]
		Power decline rate	1 (1, 0)	[78]
		Subcooling margin recovery time	1 (0, 1)	[72]

[1] The numbers indicate: total number of times used (number of times used in HSI Evaluation Studies, number of times used in Methodology Studies).

3.3 Psychometric Properties of the Human Performance Measures

Examination of the psychometric property of the human performance measures developed in the nuclear domain revealed that many of the measures have not established acceptable psychometric properties (see **Table IV**). In most cases, no evaluation of a specific psychometric property was performed at all for the measure. In some cases, evaluation was performed but the result did not meet the pre-established wide-accepted psychometric criteria. For some measures, evaluations were performed but not with the necessary detail for the research team to apply the psychometric criteria. The measures that achieved acceptable levels of reliability, validity, and sensitivity included: OPAS; Team Workload Assessment (TWA); the Halden Cooperation Scale (HCS); trust in automation scale; and the Subjective Task Complexity Scale.

Table IV. Psychometric properties of the performance measures developed in the nuclear power domain.

Instrument	Reliability		Validity				Sensit-ivity
	Internal Consistency	Inter-Rater Reliability	Content Validity	Structural Validity	Convergent Validity	Concurrent Validity	
Behavioral Performance							
OPAS	n/a	+	+	n/a	+	+	+

Instrument	Reliability		Validity				Sensit-ivity
	Internal Consistency	Inter-Rater Reliability	Content Validity	Structural Validity	Convergent Validity	Concurrent Validity	
Workload							
COCOA	n/a		+	n/a			
Activity level	n/a			n/a			
TWA	+	n/a	+				+
SA							
SACRI	n/a		+	n/a			+
Process Overview	n/a	–	+	n/a	–		+
HOPE	n/a			n/a			
IPAQ		n/a		n/a			
ASURS	n/a	-		n/a			
VISA	n/a	n/a		n/a	+/–		
Human-Human or Human-Machine Cooperation							
HCS - human-human	+	n/a	+	+		+/–	
HCS - human-automation	+	n/a	+	+		+	+
Trust in automation scale - general	+	n/a	+	+		+	+
Trust in automation scale - specific	+	n/a	+	+/–		+	–
Other Cognitive Performance Indicators							
FIR	n/a			n/a			
SAE	n/a			n/a		+	+
Perception effectiveness scale	+	n/a					
Cognition effectiveness scale	+	n/a					
Subject Task Complexity Scale	+	n/a	+	+		+	+
Metacognitive bias	n/a	n/a		n/a			
Metacognitive accuracy	n/a	n/a		n/a			
Manual responses effectiveness	+	n/a					
Vocal responses effectiveness	+	n/a					
System Performance							
APROX	n/a			n/a			
PPAS	n/a	n/a	+	n/a			
Cool-down and stabilization	n/a	n/a		n/a			
Power decline rate	n/a	n/a		n/a			
Subcooling margin recovery time	n/a	n/a		n/a		+	

+, the measure achieved acceptable level of the corresponding psychometric property in at least one study; **–**, the measure did not achieve acceptable level of the corresponding psychometric property in at least one study; n/a, the corresponding psychometric property is not applicable to the measure.

Psychometric properties of measures developed outside of NPP domain were rarely evaluated. **Table V** summarizes available evidences from the eligible studies. It should be noted that many assessments have been completed for these measures in other domains (e.g., [101; 102]) but those results were not included in this review.

Table V. Psychometric properties of the domain-generic performance measures.

Instrument	Reliability		Validity				Sensitivity
	Internal Consistency	Inter-Rater Reliability	Content Validity	Structural Validity	Convergent Validity	Concurrent Validity	
Workload							
NASA TLX	+	n/a			+	+	–
Cardiac activity - heart rate	n/a	n/a	n/a	n/a			–
Cardiac activity - heart rate variability	n/a	n/a	n/a	n/a	–	–	+/–
Eye tracking - blink rate	n/a	n/a	n/a	n/a		–	+
Eye tracking - blink duration	n/a	n/a	n/a	n/a		–	
Eye tracking - duration of fixations	n/a	n/a		n/a	+		
Eye tracking - pupil size	n/a	n/a	n/a	n/a		–	
Electroencephalogram (EEG)	n/a	n/a	n/a	n/a			–
Integrated score of physiological signals	n/a	n/a	n/a	n/a	+		
SA							
SART	–	n/a		+		–	

+, the measure achieved acceptable level of the corresponding psychometric property in at least one study; **–**, the measure did not achieve acceptable level of the corresponding psychometric property in at least one study; n/a, the corresponding psychometric property is not applicable to the measure.

4 DISCUSSION

This systematic review identified human performance measures that were used in the NPP HSI design and evaluation studies. These measures covered six dimensions of human performance: behavioral performance, workload, SA, human-human or human-machine cooperation, other cognitive performance indicators, and system performance. Available evidence of the psychometric properties of the measures were summarized and evaluated.

A number of NPP domain-specific human performance measures were identified and some of these measures showed acceptable validity, reliability, and sensitivity. NPP control room operators' work has unique sets of related tasks. For example, the operators must monitor slow, noisy, and continuous changes of large numbers of coupled parameters in a complex closed system [103]. Their monitoring tasks are characterized by actively searching for information in order to build operating context, sampling information based on the operating context, and processing meanings of information given the operating context, plant process behaviors, instrumentation, and other ongoing activities [104]. These characterizations can lead to the measurement of different components of SA which can be compared to other domains such as aviation and military command and control [105]. Well-designed domain-specific measures have the potential to provide valid and reliable information that reflects context-dependent changes in human performance. In addition, domain-specific measures can potentially be more useful than domain-generic measures in guiding design improvements as they may be designed to yield targeted meaningful diagnostic information.

This review identified gaps in the research of human performance measures that future studies should address. First, many of the domain-specific measures need additional evaluations to establish acceptable psychometric properties. As shown in **Table IV**, the psychometric properties of majority of the measures were inadequately evaluated or reported. Some measures may need further refinements to meet the acceptable psychometric criteria. However, it should be noted that some psychometric properties might be particularly difficult to establish for specific measures. For example, criterion validity was rarely established for SA measures [106-108], possibly due to (1) the need for relatively large sample size that may not be available in industrial settings and (2) different conceptual accounts of SA in different measures.

Second, more assessments and refinements are needed to establish the applicability and usefulness of the domain-generic human performance measures in the NPP domain. For example, widely used NASA TLX and SART have not showed excellent psychometric properties in NPP HSI evaluation studies. While being able to provide good temporal resolution, the physiological indicators of human performance may lack validity in naturalistic test environments, as the results can be a combination of cognitive and affective factors and physical movements [109]. Establishment of an appropriate physiological baseline is often challenging [110]. All these factors can contribute to the difficulty in establishing validity and sensitivity of physiological measures. These domain-generic measures may have to be refined in their content and the methods used for data collection and analysis in the NPP domain.

Third, the intrusiveness and utility of the measures need to be systematically evaluated. *Intrusiveness* refers to whether the administration of a measure influences the participant's performance and *utility* refer to the cost-effectiveness of a measure [9; 19]. These attributes need to be considered during measure selection, however, they are often difficult to evaluate. For example, SA Freeze probe techniques require a freeze of on-going activities for questionnaire answering or interview. It can be argued that the operators are desensitized to scenario freezes, as this is the way some training instructions are delivered; however, this might still interfere with operations and subsequently influence certain aspects of performance [111].

Fourth, there is a need to establish human performance benchmarks for the measures. This is especially relevant to HSI human factors validation activities which aim to establish that human performance while ensuring the system meets a minimum standard [6; 14]. With the available measures, it is usually unclear what this standard is [14]. In the case of HSI upgrade, one approach is to establish a human performance baseline using a set of measures with the system currently in use, and compare it to the human performance measured using the same set of measures with the new system [2]. However, this approach may be limited by study sample size – the number of available study participants in a unique plant may not provide enough power to detect human performance differences. In this case, *measurement error* reliability should be established for the human performance measures to identify the *minimal important change*, which indicates the smallest meaningful change in the scale of a measure [112].

A limitation of the current systematic review is that not all relevant studies were reviewed. As some studies were conducted by proprietary entities, they might not be made public. Limiting the review to articles that were accessible by the research team in full text and were written in English also excluded some studies that would otherwise be eligible. Another limitation was that the current review did not include psychometric evidence of the measures from other domains due to resource limitations.

5 CONCLUSIONS

As a step towards developing a guideline for human performance measure selection and use in NPP control room HSI design and evaluation, this systematic review is an environmental scan of available measures and their psychometric properties. The results indicate that future research needs to: establish acceptable level of psychometric properties for many of the domain-specific measures; assess and refine the domain-generic measures to be used in NPP domain; evaluate measures' intrusiveness and utility; and establish human performance benchmarks for the measures. In the next step, the research team will strive to integrate the best available evidence to develop best practices for how these human performance measures should be used in the evaluation of different HSI components under different study designs.

6 REFERENCES

1. WORLD NUCLEAR ASSOCIATION, "Nuclear Power in the USA," World Nuclear Association,: http://www.world-nuclear.org/information-library/country-profiles/countries-t-z/usa-nuclear-power.aspx (current as of February 25th, 2017)
2. R. L. BORING, J. C. JOE, "Baseline Evaluations to Support Control Room Modernization at Nuclear Power Plants " INL/CON-15-34403, Idaho National Laboratory (2015).
3. R. L. BORING, V. AGARWAL, J. C. JOE, J. J. PERSENSKY, "Design and validation of control room upgrades using a research simulator facility," presented at Winter Meeting of the American Nuclear Society, San Diego, CA, 2012.
4. R. L. BORING, J. C. JOE, T. A. ULRICH, R. T. LEW, "Early-stage design and evaluation for nuclear power plant control room upgrades," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, pp. 1909-1913, SAGE Publications, (2014).
5. J. C. JOE, R. L. BORING, J. J. PERSENSKY, "Commercial Utility Perspectives on Nuclear Power Plant Control Room Modernization " INL/CON-12-26175, Idaho National Laboratory (2012).
6. R. L. BORING, "Envy In V&V: An Opinon Piece on New Directions for Verification and Validation in Nuclear Power Plants," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, pp. 1746-1750, SAGE Publications Sage CA: Los Angeles, CA, (2015).
7. N. STANTON, "The discipline of human factors," in: N. STANTON Ed., *Human Factors in Nuclear Safety*, pp. 1-14, Taylor & Francis Inc., Beistol, PA, (1996).
8. C. BABER, "Human factors of human-computer interaction techniques in the control room," in: N. STANTON Ed., *Human Factors in Nuclear Safety*, pp. 51-75, Taylor & Francis Inc., Beistol, PA, (1996).
9. J. M. O'HARA, J. C. HIGGINS, S. A. FLEGER, P. A. PIERINGER, "Human Factors Engineering Program Review Model," NUREG-0711, Rev. 3, United States Nuclear Regulatory Commission (2012).
10. R. BORING, N. LAU, "Verification and Validation of Digitally Upgraded Control Rooms," INL/EXT-15-367, Idaho National Laboratory (2015).
11. L. H. IKUMA, C. KOFFSKEY, C. M. HARVEY, "A Human Factors-Based Assessment Framework for Evaluating Performance in Control Room Interface Design," *IIE Transactions on Occupational Ergonomics and Human Factors*, **2**, *3-4*, 194-206, (2014).
12. J. S. HA, P. H. SEONG, M. S. LEE, J. H. HONG, "Development of human performance measures for human factors validation in the advanced MCR of APR-1400," *IEEE Transactions on Nuclear Science*, **54**, *6*, 2687-2700, (2007).
13. ELECTRIC POWER RESEARCH INSTITUTE, "Human Factors Guidance for Control Room and Digital Human-System Interface Design and Modification: Guidelines for Planning, Specification, Design, Licensing, Implementation, Training, Operation, and Maintenance," EPRI TR-1010042, Electric Power Research Institute (2005).
14. R. BORING, J. JOE, T. ULRICH, R. LEW, "Operator Performance Metrics for Control Room Modernization: A Practical Guide for Early Design Evaluation," INL/EXT-14-31511, Idaho National Laboratory (2015).
15. C. B. TERWEE, S. D. BOT, M. R. DE BOER, D. A. VAN DER WINDT, D. L. KNOL, J. DEKKER, L. M. BOUTER, H. C. DE VET, "Quality criteria were proposed for measurement properties of health status questionnaires," *Journal of Clinical Epidemiology*, **60**, *1*, 34-42, (2007).
16. T. CLINTON-MCHARG, S. L. YOONG, F. TZELEPIS, T. REGAN, A. FIELDING, E. SKELTON, M. KINGSLAND, J. Y. OOI, L. WOLFENDEN, "Psychometric properties of implementation measures for public health and community settings and mapping of constructs against the Consolidated Framework for Implementation Research: a systematic review," *Implementation Science*, **11**, *1*, 148, (2016).
17. G. DOMINO, M. L. DOMINO, *Psychological Testing: An Introduction*, Cambridge University Press, New York, NY, (2006).
18. W. M. TROCHIM, *The Research Methods Knowledge Base*, 2nd ed, Atomic Dog Publishing, Cincinnati, OH, (2000).
19. G. ANDRESEN, A. DRØIVOLDSMO, "Human Performance Assessment: Methods and Measures," HPR-353, OECD Halden Reactor Project, Institutt for Energiteknikk (2000).
20. C. B. HOLMSTRÖM, W. R. NELSON, "Experimental evaluation of a diagnostic rule-based expert system for the nuclear industry," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, **34**, 1195-1199, (1990).
21. X. DONG, Z. LI, "A study on human redundancy in execution of computerized emergency operating procedures," *2010 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 1396-1400, (2010).
22. G. F. LIANG, J. T. LIN, S. L. HWANG, F. H. HUANG, T. C. YENN, C. C. HSU, "Evaluation and prediction of on- line maintenance workload in nuclear power plants," *Human Factors and Ergonomics in Manufacturing & Service Industries*, **19**, *1*, 64-77, (2009).
23. M.-H. HSIEH, S.-L. HWANG, K.-H. LIU, S.-F. M. LIANG, C.-F. CHUANG, "A decision support system for identifying abnormal operating procedures in a nuclear power plant," *Nuclear Engineering and Design*, **249**, 413-418, (2012).

24. M.-C. HSIEH, M.-C. CHIU, S.-L. HWANG, "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, (2015).
25. F.-H. HUANG, Y.-L. LEE, S.-L. HWANG, T.-C. YENN, Y.-C. YU, C.-C. HSU, H.-W. HUANG, "Experimental evaluation of human–system interaction on alarm design," *Nuclear Engineering and Design*, **237**, *3*, 308-315, (2007).
26. F.-H. HUANG, S.-L. HWANG, "Experimental studies of computerized procedures and team size in nuclear power plant operations," *Nuclear Engineering and Design*, **239**, *2*, 373-380, (2009).
27. J. H. KIM, P. H. SEONG, "The effect of information types on diagnostic strategies in the information aid," *Reliability Engineering & System Safety*, **92**, *2*, 171-186, (2007).
28. Y.-L. LEE, S.-L. HWANG, E. M.-Y. WANG, "Reducing cognitive workload of a computer-based procedure system," *International journal of human-computer studies*, **63**, *6*, 587-606, (2005).
29. J. PARK, W. JUNG, K. JUNG, "The effect of two complexity factors on the performance of emergency tasks — an experimental verification," *Reliability Engineering & System Safety*, **93**, *2*, 350-362, (2008).
30. S.-L. HWANG, J.-T. LIN, G.-F. LIANG, Y.-J. YAU, T.-C. YENN, C.-C. HSU, "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, (2008).
31. S. XU, F. SONG, Z. LI, Q. ZHAO, W. LUO, X. HE, G. SALVENDY, "An ergonomics study of computerized emergency operating procedures: presentation style, task complexity, and training level," *Reliability Engineering & System Safety*, **93**, *10*, 1500-1511, (2008).
32. X. DING, Z. LI, X. DONG, Q. GAO, F. SONG, Q. WANG, "Effects of information organization and presentation on human performance in simulated main control room procedure tasks," *Human Factors and Ergonomics in Manufacturing & Service Industries*, **25**, *6*, 713-723, (2015).
33. Y.-T. JOU, T.-C. YENN, C. J. LIN, C.-W. YANG, C.-C. CHIANG, "Evaluation of operators' mental workload of human–system interface automation in the advanced nuclear power plants," *Nuclear Engineering and Design*, **239**, *11*, 2537-2542, (2009).
34. Y.-T. JOU, T.-C. YENN, C. J. LIN, C.-W. YANG, S.-F. LIN, "Evaluation of mental workload in automation design for a main control room task," *Proceedings of the 2009 International Conference on Networking, Sensing and Control*, 313-317, (2009).
35. H. B. YIM, I. KIM, P. H. SEONG, "An Abstraction Hierarchy based mobile PC display design in NPP maintenance considering the level of expertise," *Nuclear Engineering and Design*, **241**, *5*, 1881-1888, (2011).
36. S. K. KIM, S. M. SUH, G. S. JANG, S. K. HONG, J. C. PARK, "Empirical research on an ecological interface design for improving situation awareness of operators in an advanced control room," *Nuclear Engineering and Design*, **253**, 226-237, (2012).
37. S. H. LAU, J. A. BERNARD, K. S. KWOK, D. D. LANNING, "Experimental evaluation of predictive information as an operator aid in the control of research reactor power," *Proceedings of the 1988 American Control Conference*, 214-220, (1988).
38. S. A. CONVERSE, "Operating procedures: do they reduce operator errors?," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, **38**, 205-209, (1994).
39. Q. GAO, Y. WANG, F. SONG, Z. LI, X. DONG, "Mental workload measurement for emergency operating procedures in digital nuclear power plants," *Ergonomics*, **56**, *7*, 1070-1085, (2013).
40. M. HILDEBRANDT, M. H. R. EITRHEIM, "A micro-task method for assessing performance effects of innovative interface elements," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, **59**, 1759-1763, (2015).
41. G. JIANG, P. GU, Y. SUN, "A Study About Human Workload Forecasting Method of Nuclear Power Plant," *18th International Conference on Nuclear Engineering*, pp. 743-748, American Society of Mechanical Engineers, Xi'an, China, (2010).
42. E. HOLLNAGEL, A. B. MIBERG, "Human-Centred Automation: an Explorative Study (HWR-595)," OECD Halden Reactor Project, Institutt for Energiteknikk (1999).
43. M.-C. HSIEH, M.-C. CHIU, S.-L. HWANG, "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, (2014).
44. A. B. M. SKJERVE, S. STRAND, R. SAARNI, G. SKRAANING JR, "The Influence of Automation Malfunctions and Interface Design on Operator Performance. The HCA-2001 Experiment (HWR-686)," OECD Halden Reactor Project, Institutt for Energiteknikk (2002).
45. P. V. R. CARVALHO, J. O. GOMES, M. R. S. BORGES, "Human centered design for nuclear power plant control room modernization," *CEUR Proceedings of the 4th Workshop on Human Centered Processes*, 25-33, (2011).
46. J. LAARNI, H. KOSKINEN, L. SALO, L. NORROS, A. BRASETH, V. NURMILAUKAS, "Evaluation of the Fortum IRD pilot.," *Proceedings of the Sixth American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies NPIC&HMIT*, (2009).
47. G. ANDRESEN, H. SVENGREN, J. O. HEIMDAL, S. NILSEN, J.-E. HULSUND, R. BISIO, X. DEBROISE, "Procedure Automation: the Effect of Automated Procedure Execution on Situation Awareness and Human Performance (HWR-759)," OECD Halden Reactor Project, Institutt for Energiteknikk (2004).
48. S. STRAND, "Trust and Automation: the Influence of Automation Malfunctions and System Feedback on Operator Trust (HWR-643)," OECD Halden Reactor Project, Institutt for Energiteknikk (2001).

49. C. J. LIN, T. L. HSIEH, P. J. TSAI, C. W. YANG, T. C. YENN, "Development of a team workload assessment technique for the main control room of advanced nuclear power plants," *Human Factors and Ergonomics in Manufacturing & Service Industries*, **21**, *4*, 397-411, (2011).
50. K. J. VICENTE, N. MORAY, J. D. LEE, J. R. HURECON, B. G. JONES, R. BROCK, T. DJEMIL, "Evaluation of a Rankine cycle display for nuclear power plant monitoring and diagnosis," *Human Factors*, **38**, *3*, 506-521, (1996).
51. E. M. ROTH, L. LIN, V. M. THOMAS, S. KERCH, S. J. KENNEY, N. SUGIBAYASHI, "Supporting situation awareness of individuals and teams using group view displays," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, **42**, 244-248, (1998).
52. H.-C. LEE, I.-S. OH, B.-S. SIM, I.-S. KOO, J.-T. KIM, K.-Y. LEE, J.-K. PARK, "Operator Performance Comparison of two VDT-based Alarm Systems," *Proceedings from Specialists Meeting on Human Performance in Operational Events*, 283-287, (1998).
53. S. S. CHOI, S. H. CHANG, D. H. LEE, "Automating strategies of emergency operation for optimal shutdown in pressurized water reactors," *IEEE Transactions on Nuclear Science*, **45**, *1*, 17-29, (1998).
54. S. J. LEE, P. H. SEONG, "Experimental Effect Estimation of an Integrated Decision Support System to Aid Operator's Cognitive Activities for Nuclear Power Plants," in: J. A. JACKO Ed., *Human-Computer Interaction. HCI Applications and Services: 12th International Conference, HCI International 2007, Beijing, China, July 22-27, 2007, Proceedings, Part IV*, pp. 620-628, Springer Berlin Heidelberg, Berlin, Heidelberg, (2007).
55. S. J. LEE, P. H. SEONG, "Experimental investigation into the effects of decision support systems on operator performance," *Journal of nuclear science and technology*, **46**, *12*, 1178-1187, (2009).
56. C. J. LIN, T.-L. HSIEH, C.-W. YANG, R.-J. HUANG, "The impact of computer-based procedures on team performance, communication, and situation awareness," *International Journal of Industrial Ergonomics*, **51**, 21-29, (2016).
57. F.-H. HUANG, S.-L. HWANG, T.-C. YENN, Y.-C. YU, C.-C. HSU, H.-W. HUANG, "Evaluation and comparison of alarm reset modes in advanced control room of nuclear power plants," *Safety science*, **44**, *10*, 935-946, (2006).
58. N. LAU, G. A. JAMIESON, G. SKRAANING JR, C. M. BURNS, "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, (2008).
59. N. LAU, G. SKRAANING, G. A. JAMIESON, C. M. BURNS, "Enhancing operator task performance during monitoring for unanticipated events through ecological interface design," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, **52**, 448-452, (2008).
60. N. LAU, G. SKRAANING JR, G. A. JAMIESON, "Metacognition in nuclear process control," *Proceedings of the 17th Triennial World Congress on Ergonomics*, (2009).
61. G. SKRAANING JR, "The Operator Performance Assessment System (OPAS) (HWR-538)," OECD Halden Reactor Project, Institutt for Energiteknikk (1998).
62. M. DEMAS, N. LAU, C. ELKS, "Advancing human performance assessment capabilities for integrated system validation—A human-in-the-loop experiment," *Proceedings of the 9th American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation & Control and Human-Machine Interface Technologies (NPIC & HMIT)*, 1051-1064, (2015).
63. N. LAU, G. A. JAMIESON, G. SKRAANING JR, "Inter-rater reliability of expert-based performance measures," *Proceedings of the 8th American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation & Control and Human-Machine Interface Technologies (NPIC & HMIT)*, 1974-1982, (2012).
64. N. LAU, G. A. JAMIESON, G. SKRAANING JR, "Empirical evaluation of the Process Overview Measure for assessing situation awareness in process plants," *Ergonomics*, 1-16, (2016).
65. N. LAU, G. SKRAANING JR, M. H. EITRHEIM, T. KARLSSON, C. NIHLWING, G. A. JAMIESON, "Situation awareness in monitoring nuclear power plants: The Process Overview concept and measure," *Halden, Norway: OECD Halden Reactor Project*, (2011).
66. A. B. M. SKJERVE, "The Halden Co-operation Scale. Human-Automation Co-operation in Control Room Settings (HWR-685)," OECD Halden Reactor Project, Institutt for Energiteknikk (2002).
67. J. O'HARA, W. BROWN, B. HALLBERT, G. SKRANING, J. PERSENSKY, J. WACHTEL, "The Effects of Alarm Display, Processing, and Availability on Crew Performance (NUREG/CR-6691, BNL-NUREG-52600)," U.S. Nuclear Regulatory Commission (2000).
68. A. SEBOK, "Team performance in process control: influences of interface design and staffing levels," *Ergonomics*, **43**, *8*, 1210-1236, (2000).
69. C. HUMBEL, D. TASSET, "Advanced human-system interfaces: Insight of a validation process in a control room of a Swiss nuclear power plant," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, **44**, 185-188, (2000).
70. L. NORROS, M. NUUTINEN, "Performance-based usability evaluation of a safety information and alarm system," *International Journal of Human-Computer Studies*, **63**, *3*, 328-361, (2005).
71. S. STRAND, H. SVENGREN, C. NIHLWING, L. KRISTIANSEN, G. ANDRESEN, B. MEYER, "Task-Based Displays — Prototype Extensions and the Second User Test (HWR-841)," OECD Halden Reactor Project, Institutt for Energiteknikk (2007).
72. I. JANG, J. PARK, P. SEONG, "An empirical study on the relationships between functional performance measure and task performance measure in NPP MCR," *Annals of Nuclear Energy*, **42**, 96-103, (2012).

73. P. Ø. BRAARUD, "Subjective Task Complexity in the Control Room (HWR-621)," OECD Halden Reactor Project, Institutt for Energiteknikk (2000).
74. Y.-T. JOU, T.-C. YENN, C. J. LIN, W.-S. TSAI, T.-L. HSIEH, "The research on extracting the information of human errors in the main control room of nuclear power plants by using Performance Evaluation Matrix," *Safety science*, **49**, *2*, 236-242, (2011).
75. C.-W. YANG, L.-C. YANG, T.-C. CHENG, Y.-T. JOU, S.-W. CHIOU, "Assessing mental workload and situation awareness in the evaluation of computerized procedures in the main control room," *Nuclear Engineering and Design*, **250**, 713-719, (2012).
76. C. J. LIN, T.-C. YENN, Y.-T. JOU, T.-L. HSIEH, C.-W. YANG, "Analyzing the staffing and workload in the main control room of the advanced nuclear power plant from the human information processing perspective," *Safety science*, **57**, 161-168, (2013).
77. C. H. HA, J. H. KIM, S. J. LEE, P. H. SEONG, "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, (2006).
78. Y. T. JOU, T. C. YENN, L. C. YANG, "Investigation of automation deployment in the main control room of nuclear power plants by using adaptive automation," *Human Factors and Ergonomics in Manufacturing & Service Industries*, **21**, *4*, 350-360, (2011).
79. R. BORING, R. LEW, T. ULRICH, J. JOE, "Light Water Reactor Sustainability Program Operator Performance Metrics for Control Room Modernization: A Practical Guide for Early Design Evaluation (INL/EXT-14-31511)," Idaho National Laboratory (2014).
80. C. J. LIN, T.-C. YENN, C.-W. YANG, "Automation design in advanced control rooms of the modernized nuclear power plants," *Safety science*, **48**, *1*, 63-71, (2010).
81. T.-I. JANG, J.-K. PARK, D.-H. KIM, J.-W. LEE, Y.-H. LEE, "An experimental evaluation of the flowchart type Operating Procedure for Nuclear Power Plants," *Proceedings of the 2007 IEEE 8th Human Factors and Power Plants and HPRCT 13th Annual Meeting*, 152-155, (2007).
82. J. E. MERCADO, L. REINERMAN-JONES, D. BARBER, R. LEIS, "Investigating workload measures in the nuclear domain," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, **58**, 205-209, (2014).
83. S.-L. HWANG, Y.-J. YAU, Y.-T. LIN, J. H. CHEN, T.-H. HUANG, T.-C. YENN, C.-C. HSU, "A mental workload predicator model for the design of pre alarm systems," *Engineering Psychology and Cognitive Ergonomics: 7th International Conference, EPCE 2007, Held as Part of HCI International 2007, Beijing, China, July 22-27, 2007. Proceedings*, pp. 316-323, Springer Berlin Heidelberg, Berlin, Heidelberg, (2007).
84. J. PARK, W. JUNG, "A study on the validity of a task complexity measure for emergency operating procedures of nuclear power plants: Comparing with a subjective workload," *IEEE Transactions on Nuclear Science*, **53**, *5*, 2962-2970, (2006).
85. A. DRØEIVOLDSMO, G. SKRAANING JR, M. SVERRBO, J. DALEN, T. GRIMSTAD, G. ANDRESEN, "Continuous Measures of Situation Awareness and Workload (HWR-539)," OECD Halden Reactor Project, Institutt for Energiteknikk (1998).
86. P. Ø. BRAARUD, H. BRENDRYEN, "Task Demand, Task Management, and Teamwork (HWR-657)," OECD Halden Reactor Project, Institutt for Energiteknikk (2001).
87. K. CHEN, Z. LI, "Evaluation of human-system interfaces with different information organization using an eye tracker," *International Conference on Cross-Cultural Design*, 288-295, (2013).
88. Y. KIM, W. JUNG, S. KIM, "Empirical investigation of workloads of operators in advanced control rooms," *Journal of Nuclear Science and Technology*, **51**, *6*, 744-751, (2014).
89. S. KIM, Y. KIM, W. JUNG, "Operator's cognitive, communicative and operative activities based workload measurement of advanced main control room," *Annals of Nuclear Energy*, **72**, 120-129, (2014).
90. E. ROTH, J. O'HARA, "Integrating Digital and Conventional Human-System Interfaces: Lessons Learned from a Control Room Modernization Program (NUREG/CR-6749, BNL-NUREG-52638)," U.S. Nuclear Regulatory Commission (2002).
91. K. J. VICENTE, R. J. MUMAW, E. M. ROTH, "More about operator monitoring under normal operations: The role of workload regulation and the impact of control room technology," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, **42**, 229-233, (1998).
92. D. N. HOGG, K. FOLLES, F. S. VOLDEN, B. TORRALBA, "SACRI: A measure of situation awareness for use in the evaluation of nuclear power plant control room systems providing information about the current process state," *Specialists' Meeting on Advanced Information Methods and Artificial Intelligence in Nuclear Power Plant Control Rooms*, Halden, Norway, (1995).
93. D. N. HOGG, K. FOLLES⊘, F. STRAND-VOLDEN, B. TORRALBA, "Development of a situation awareness measure to evaluate advanced alarm systems in nuclear power plant control rooms," *Ergonomics*, **38**, *11*, 2394-2413, (1995).
94. C. M. BURNS, G. A. JAMIESON, G. SKRAANING, N. LAU, J. KWOK, "Supporting situation awareness through ecological interface design," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, **51**, 205-209, (2007).
95. C. M. BURNS, G. SKRAANING, G. A. JAMIESON, N. LAU, J. KWOK, R. WELCH, G. ANDRESEN, "Evaluation of ecological interface design for nuclear process control: situation awareness effects," *Human Factors*, **50**, *4*, 663-679, (2008).
96. N. LAU, G. A. JAMIESON, G. SKRAANING JR, "Inter-rater reliability of query/probe-based techniques for measuring situation awareness," *Ergonomics*, **57**, *7*, 959-972, (2014).

97. M. R. ENDSLEY, S. J. SELCON, T. D. HARDIMAN, D. G. CROFT, "A comparative analysis of SAGAT and SART for evaluations of situation awareness," *Proceedings of the human factors and ergonomics society annual meeting*, pp. 82-86, SAGE Publications Sage CA: Los Angeles, CA, (1998).
98. J. S. HA, P. H. SEONG, "Attentional-resource effectiveness measures in monitoring and detection tasks in nuclear power plants," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, **40**, *5*, 993-1008, (2010).
99. J. S. HA, P. H. SEONG, "A human–machine interface evaluation method: A difficulty evaluation method in information searching (DEMIS)," *Reliability Engineering & System Safety*, **94**, *10*, 1557-1567, (2009).
100. M. J. MORACHO, "Plant Performance Assessment System (PPAS) for Crew Performance Evaluation. Lessons Learned from an Alarm System Study Conducted in HAMMLAB (HWR-504)," OECD Halden Reactor Project, Institutt for Energiteknikk (1998).
101. P. M. SALMON, N. A. STANTON, G. H. WALKER, D. JENKINS, D. LADVA, L. RAFFERTY, M. YOUNG, "Measuring Situation Awareness in complex systems: Comparison of measures study," *International Journal of Industrial Ergonomics*, **39**, *3*, 490-500, (2009).
102. S. G. HART, "NASA-task load index (NASA-TLX); 20 years later," *Proceedings of the human factors and ergonomics society annual meeting*, pp. 904-908, Sage Publications Sage CA: Los Angeles, CA, (2006).
103. N. LAU, G. A. JAMIESON, G. SKRAANING JR, "Situation awareness in process control: a fresh look," *Proceedings of the 8th American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation & Control and Human-Machine Interface Technologies (NPIC & HMIT), San Diego, CA, USA*, (2012).
104. N. LAU, G. A. JAMIESON, G. SKRAANING JR, "Situation awareness acquired from monitoring process plants–the Process Overview concept and measure," *Ergonomics*, **59**, *7*, 976-988, (2016).
105. N. LAU, G. A. JAMIESON, G. SKRAANING, "Distinguishing three accounts of situation awareness based on their domains of origin," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, pp. 220-224, SAGE Publications, (2013).
106. D. G. JONES, M. R. ENDSLEY, "Use of real-time probes for measuring situation awareness," *The International Journal of Aviation Psychology*, **14**, *4*, 343-367, (2004).
107. N. LAU, G. A. JAMIESON, G. SKRAANING JR, "Empirical evaluation of the Process Overview Measure for assessing situation awareness in process plants," *Ergonomics*, **59**, *3*, 393-408, (2016).
108. K. O'BRIEN, D. O'HARE, "Situational awareness ability and cognitive skills training in a complex real-world task," *Ergonomics*, **50**, *7*, 1064-1091, (2007).
109. W. BOUCSEIN, R. W. BACKS, "Engineering psychophysiology as a discipline: Historical and theoretical aspects," in: R. W. BACKS, W. BOUCSEIN Eds., *Engineering psychophysiology. Issues and applications*, pp. 3-30, Lawrence Erlbaum Associates, Inc., Mahwah, NJ, (2000).
110. A. PARCHMENT, R. W. WOHLEBER, L. REINERMAN-JONES, "Psychophysiological Baseline Methods and Usage," *International Conference on Augmented Cognition*, pp. 361-371, Springer, (2016).
111. M. M. CHATZIMICHAILIDOU, A. PROTOPAPAS, I. M. DOKAS, "Seven Issues on Distributed Situation Awareness Measurement in Complex Socio-technical Systems," *Complex Systems Design & Management*, pp. 105-117, Springer, (2015).
112. C. B. TERWEE, L. D. ROORDA, D. L. KNOL, M. R. DE BOER, H. C. DE VET, "Linking measurement error to minimal important change of patient-reported outcomes," *Journal of clinical epidemiology*, **62**, *10*, 1062-1067, (2009).