Human Factors in Design, Engineering, and Computing, Vol. 159, 2024, 948–957
https://doi.org/10.54941/ahfe1005661

Revisiting the Brief Nuclear Usability Measure: A Preliminary Evaluation of Its Validity and Reliability Using Licensed Operators

Casey Kovesdi

Idaho National Laboratory, Idaho Falls, ID 83415, USA

ABSTRACT

Nuclear power has a critical role in addressing global demands for electricity. The continued operation of existing plants and the development of new advanced reactors will be needed to support these electricity needs. In ensuring safe and reliable operation of these plants, human factors engineering has an important role to address the design of human-system interfaces, procedures, and training. Human factors engineering must rely on valid, reliable, and practical tools to evaluate important constructs such as workload and situation awareness. This work presents an exploratory evaluation of the validity and reliability for a prospective tool that evaluates these constructs, using real-world data collected from three separate and independent human factors engineering studies with licensed operators. The intent of this work is to provide an early assessment of the utility of this measure as a practical tool that can be used by human factors practitioners to evaluate situation awareness and workload for prospective human factors tests and evaluations for both modernization efforts and advanced reactor development.

Keywords: Nuclear human factors engineering, Workload, Situation awareness, Survey design

INTRODUCTION

Global energy consumption is expected to increase through 2050 (EIA, 2023). Population grown, increased manufacturing, and higher living standards are cited as key drivers to pushing energy consumption beyond energy efficiency advances. In order to circumvent such grim projections, the role of nuclear electricity generation has a pivotal role in providing carbon-free electricity generation across the world. Within the United States, there have been thrusts in extending the operational lifespan of the existing light water reactor fleet through significantly modernizing these existing plants with digital technologies that reduce their operations and maintenance cost. Additionally, despite some setbacks, development and deployment of advanced reactor technologies are continuing to move forward from both developer and regulatory standpoints.

Across both strategies of deploying reactor technology to address these global challenges, the role of human factors engineering is crucial to ensuring timely completion of major modernization efforts or advanced reactor deployment by addressing human and technology integration challenges. Such challenges range from effective allocation of function with digital technologies and automation to the design of novel human-system interfaces that support effective monitoring and control of advanced reactors. Two important human factors constructs that are relevant for human factors tests and evaluation of these advanced technologies entails situation awareness and workload (e.g., NUREG/CR-7190, 2015). For instance, these constructs are referenced in existing regulatory review guidance (e.g., NUREG-0711) as vital to assess during integrated system validation and other testing and evaluation activities.

Indeed, a multi-method and multi-measure approach is generally most appropriate for assessing situation awareness and workload as these constructs are not directly observable (Matthews & Reinerman-Jones, 2017). A suite of objective and subjective measures is typically a "gold standard" in evaluating situation awareness and workload. Though, the use of self-report through standardized survey instruments offers a practical way of collecting such data, particularly in the "real world." For instance, when testing licensed operators, availability and time for data collection can be significantly limited (Kovesdi and Joe, 2019). Therefore, approaches to streamline survey instruments that can adequately assess situation awareness and workload have been explored. Kovesdi and Joe (2019) developed an abbreviated survey instrument, the Brief Nuclear Usability Measure, derived from reviewing the National Aeronautics and Space Administration Raw Task Load Index (NASA-TLX; Hart and Staveland, 1988) Single Ease Question (SEQ; Sauro and Lewis, 2016), and Situation Awareness Rating Technique (SART; Taylor, 1990).

This work presents an exploratory evaluation of the validity and reliability of the Brief Nuclear Usability Measure (BNUM) using real-world data collected from three separate and independent human factors engineering studies that utilized licensed operators during operator-in-the-loop studies. The intent of this work is to provide an early assessment of the utility of this measure as a practical tool that can be used by human factors practitioners to evaluate situation awareness and workload for prospective human factors tests and evaluations for both modernization efforts and advanced reactor development. This paper has three key sections.

The first section provides a brief overview of BNUM and provides context into the motives in developing the survey instrument. The next section presents preliminary analysis results, highlighting its correlation structure to the referred human factors engineering survey instruments commonly used as a battery to evaluate perceived situation awareness and workload from three independent human factors engineering studies that included licensed operators from three separate nuclear power plants in the United States. Finally, this paper provides a conclusion to this early assessment of BNUM and offers next steps in validating the survey instrument as a viable tool to assess perceived situation awareness and workload.

BACKGROUND

Drivers for Developing the Brief Nuclear Usability Measure

The primary motivation for developing the BNUM was to provide a *shortened practical survey instrument* that can be used to evaluate perceived situation awareness and workload in full scope, full scale human factors engineering tests and evaluations that require licensed operators (Kovesdi & Joe, 2019). Because the time available for licensed operators can be severely limited and the scope of such testing and evaluation activities are complex in nature (i.e., demanding multiple scenarios, testing and discussion activities, etc.), there is a need to maximize the time spent with licensed operators for purposes beyond collecting self-report survey data. Thus, BNUM was developed to address this need.

A second objective of the BNUM survey was to also provide *diagnostic criteria* that can be used in the review of specific responses of perceived situation awareness and workload as part of identifying attributes with the scenario, human-system interfaces, procedures, training, or artifacts that influenced the responses. This latter objective is meant to aid in the facilitation of semi-structured discussions that are commonly performed after specific scenario runs.

Characteristics of the Brief Nuclear Usability Measure

The BNUM survey was developed through examining the psychometric properties of three combined survey instruments used to evaluated perceived situation awareness and workload.

1. The SART – Situation Awareness
2. The NASA-TLX – Workload
3. SEQ – Perceived Difficulty and Workload

An in-depth description of these survey instruments go beyond the scope of this paper. Further, a discussion of whether the SART is a valid measure of situation awareness as opposed to objective measures like the Situation Awareness Global Assessment Technique (SAGAT) goes beyond this work here. Though, it should be noted that these survey instruments are recognized across the nuclear human factors community as common tools for assessing situation awareness and workload (Braarud, 2021).

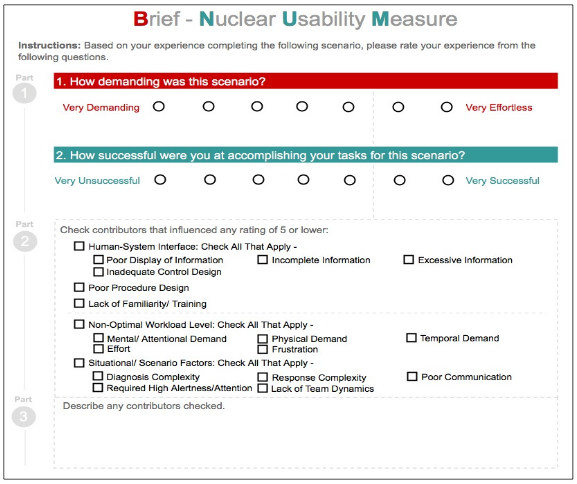
When the three survey instruments are combined, they include several items, each with their own rating scale and directionality of meaning. For instance, the NASA-TLX is commonly scored on a 20-point rating scale to which higher ratings refer to greater perceived workload. The SEQ is scored on a 7-point rating scale to which lower ratings refer to greater perceived workload. Finally, the SART has a series of three types of rating questions that each refer to a dimension of situation awareness: attentional supply, attentional demand, and understanding of the situation. Their directionality to their rating influences situation awareness differently. Ratings for questions related to attentional supply and understanding of the situation have a positive influence on perceived situation awareness whereas ratings for questions related to attentional demand have a negative influence.

Beyond the drivers previously highlighted in this paper, an observation made here is that these differences between the inherent qualities of each survey instrument can create added complexity in interpreting each question. In fact, the author here has antidotally observed expressed difficulties and frustrations by licensed operators when completing these survey instruments due to this very reason. Thus, the BNUM survey was designed so that the rating directionality is consistent for aided interpretation.

The results from an exploratory principal component analysis (PCA) from the original work by Kovesdi and Joe (2019) is presented in Table 1. The table presents the factor load values of the PCA for each of the items across the NASA-TLX, SEQ, and SART to the two principle components that ultimately informed the development of BNUM. The results coming from the PCA informed the development of BNUM (Figure 1).

Table 1. Factor loading values using PCA (Kovesdi & Joe, 2019).

NASA-TLX, SEQ, and SART Items		PCA Factor Loading Values for BNUM	
NASA-TLX	Mental workload	Workload	$\lambda = 0.74$
	Physical workload	Workload	$\lambda = 0.57$
	Effort	Workload	$\lambda = 0.79$
	Temporal workload	Workload	$\lambda = 0.79$
	Performance	Situation awareness	$\lambda = -0.56$
	Frustration	Workload	$\lambda = 0.38$
SEQ	Perceived difficulty	Workload	$\lambda = -0.62$
SART	Attentional demand Q1	Workload	$\lambda = 0.72$
	Attentional demand Q2	Workload	$\lambda = 0.80$
	Attentional demand Q3	Workload	$\lambda = 0.75$
	Attentional supply Q1	Situation awareness	$\lambda = 0.92$
	Attentional supply Q2	Situation awareness	$\lambda = 0.84$
	Attentional supply Q3	Situation awareness	$\lambda = 0.74$
	Attentional supply Q4	Situation awareness	$\lambda = 0.76$
	Understanding Q1	Situation awareness	$\lambda = 0.80$
	Understanding Q2	Situation awareness	$\lambda = 0.71$


Brief - Nuclear Usability Measure

Instructions: Based on your experience completing the following scenario, please rate your experience from the following questions.

Part 1

1. How demanding was this scenario?

Very Demanding ○ ○ ○ ○ ○ ○ ○ Very Effortless

2. How successful were you at accomplishing your tasks for this scenario?

Very Unsuccessful ○ ○ ○ ○ ○ ○ ○ Very Successful

Part 2

Check contributors that influenced any rating of 5 or lower:

☐ Human-System Interface: Check All That Apply -
- ☐ Poor Display of Information ☐ Incomplete Information ☐ Excessive Information
- ☐ Inadequate Control Design

☐ Poor Procedure Design

☐ Lack of Familiarity/ Training

☐ Non-Optimal Workload Level: Check All That Apply -
- ☐ Mental/ Attentional Demand ☐ Physical Demand ☐ Temporal Demand
- ☐ Effort ☐ Frustration

☐ Situational/ Scenario Factors: Check All That Apply -
- ☐ Diagnosis Complexity ☐ Response Complexity ☐ Poor Communication
- ☐ Required High Alertness/Attention ☐ Lack of Team Dynamics

Part 3

Describe any contributors checked.

Figure 1: Copy of the BNUM survey (adapted from Kovesdi & Joe, 2019).

As seen in Figure 1, there are two primary questions for BNUM. This is presented in Part 1 of the survey. Both questions are answered using a 7-point rating scale in which higher ratings reflect improved performance denoted as reduced workload (Question 1) and increased perceived situation awareness (Question 2). For lower ratings of either question, Part 2 of BNUM provides a means for identifying contributors to the low rating. Here, there are several checkboxes grouped by contributors related to 1) the HSI, 2) the procedures, 3) training, 4) specific workload considerations, or 5) situational considerations. One or more contributors can be checked. Finally, Part 3 provides a mechanism for providing additional information around the contributors or adding to other contributors not listed in Part 2 (e.g., study artifacts).

The analysis of responses from the BNUM is generally to be interpreted by each of the two individual rating scales. For instance, perceived workload can be analyzed through descriptive and inferential statistical analysis using Question 1 whereas perceived situation awareness can be done using Question 2, respectively.

CONTEXT OF USE FOR BNUM

The main use case for applying BNUM is to evaluate perceived workload and situation awareness within the context of human factors engineering tests and evaluations that require licensed operators and the use of a full scope, full scale testbed (Joe & Kovesdi, 2021; e.g., Figure 2).

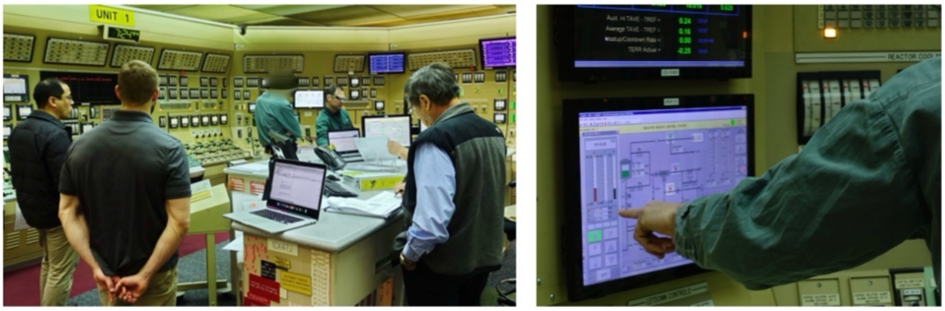


Figure 2: Photographs of licensed operators performing operational tasks in a full scope, full scale testbed (Joe & Kovesdi, 2021).

Such activities often entails a multidisciplinary team, comprised of a combination of subject matter experts in human factors engineering, instrumentation and controls (I&C) engineering, operations, and training at a minimum. The testbed in use can vary from a glasstop simulator to the use of the qualified on-site training simulator used for operator training and qualifications, such as seen in Figure 2. In either case, a general workflow is to have a set of scenarios that demonstrate key tasks, plant casualties, and interactions with the human-system interfaces of interests in a naturalistic and integrative manner.

Licensed operators perform their tasks in these scenarios without interruption while human factors engineers collect observational data that may be either exploratory in nature or evaluative using established acceptance criteria, depending on the nature and scope of the test and evaluation. Once a scenario is complete, the human factors engineers will administer the post-scenario survey(s), in which BNUM is administered. The licensed operators will complete the surveys, including BNUM, independently before concluding with a detailed debrief and post-scenario discussion facilitated by the human factors engineer and training or operational subject matter expert(s). The responses from Parts 1, 2, and 3 of BNUM can be used as a facilitation guide to query the operators' experiences when performing their tasks from the scenario. The results from BNUM are generally used to support formative analysis of identifying potential design issues with candidate human-system interfaces, procedures, or training using these parts of the survey. The ratings themselves can also be used to compare performance between different conditions, such as between scenarios or between different design options.

PRELIMINARY ANALYSIS RESULTS AND DISCUSSION

Since the development of BNUM, it has been used in combination with NASA-TLX, SEQ, and SART for several human factors engineering tests and evaluations that support 'real world' engineering efforts with nuclear power plant utilities who were undergoing significant digital modifications in their main control room and I&C systems that command and control the plant.

These modifications spanned undergoing significant modifications to the human-system interfaces, procedures, and training to which these nuclear power plants are licensed to. The span of these modifications range from replacing existing analog indications and controls with advanced digital human-system interfaces, and altering the concept of operations from performing operational tasks from walking up to the boards and physically manipulating controls to a seated operation at designated workstations that provide indications and controls from digital interfaces.

SCOPE OF PRELIMINARY ANALYSIS

Specifically, three subsequent studies, independent from the studies used to development BNUM, were used to preliminarily analysis the validity and reliability of BNUM to the original set of survey instruments: NASA-TLX, SEQ, and SART. Each of these studies were based on three separate nuclear power plants and all three were performed using a glasstop testbed, such as Idaho National Laboratory's Human-System Simulation Laboratory (Figure 3).

Figure 3: Photograph of Idaho National Laboratory's Human-System Simulation Laboratory.

Each of the three studies contained one nuclear power plant operator crew from each plant, respectively. The crew contained two licensed reactor operators (i.e., one responsible for reactor control and one responsible for balance of plant operation), one licensed senior reactor operator who supervises the crew, and a senior technical advisor who is responsible for providing a high-level independent assessment of the plant during casualty situations (e.g., during a steam generator tube rupture). A total of ten ($N = 10$) licensed operators of some combination of crew role was used in this preliminary analysis.

Despite these studies being completely independent of each other, the overall study execution flow followed a similar approach: 1) the scenario was performed, 2) surveys were administered and completed by licensed operators independently, and 3) a post-scenario debrief/ discussion was facilitated by a human factors engineer. Further, there were no modifications made to the surveys administered, and all surveys were administered via pen and paper.

Preliminary Analysis and Results

Internal consistency (reliability) was examined using Cronbach's Alpha across the survey items using the *psych* package in R (Revelle, 2023). Items intended to map to each construct (i.e., workload and situation awareness as seen in Table 1) were analyzed together. The items intended to be associated with workload showed strong internal consistency, $\alpha = 0.79$. There was notably weaker internal consistency identified for items intended to be associated with perceived situation awareness, $\alpha = 0.48$. The relationship between each associated item was examined using a series of Pearson correlations from R as an exploratory measure of validity. The correlations are visualized using R *ggcorrplot* package (Kassambara, 2023), in Figure 4.

Correlation coefficients that are statistically significant ($p < .05$) are colored and labeled and nonsignificant coefficients are left white without labels. Further, positive coefficients are orange whereas negative coefficients are purple. The coefficients of interest as they map to each BNUM question are highlighted in red and teal. The red and teal colors are also reflected in Table 1 and Figure 1.

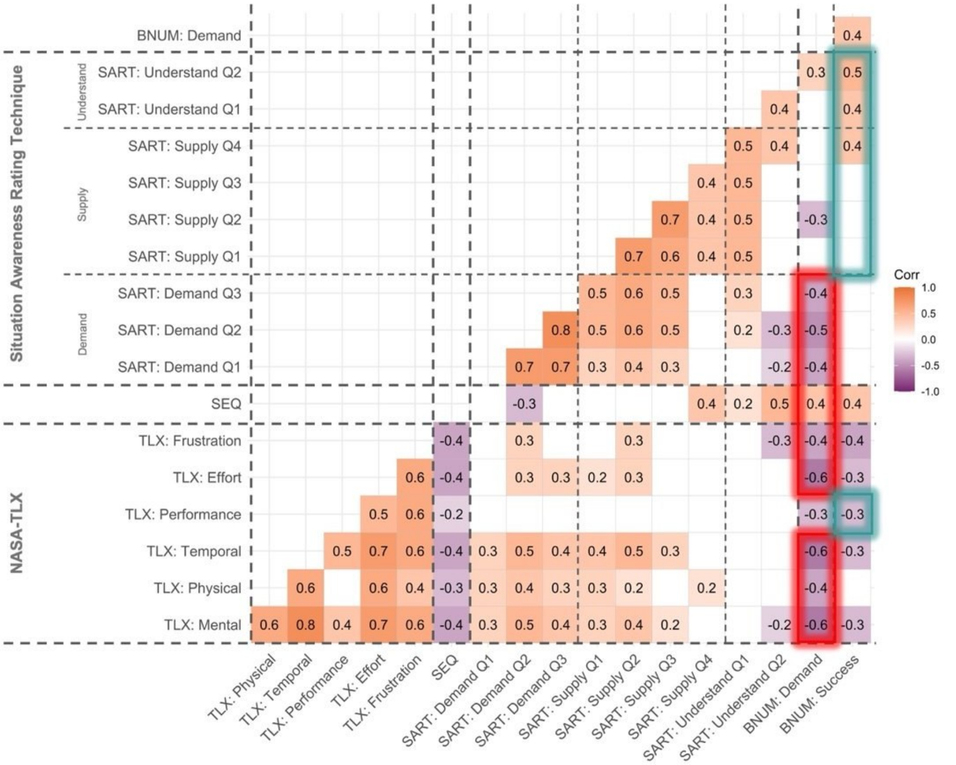


Figure 4: Pearson correlation matrix comparing NASA-TLX, SEQ, SART, and BNUM.

DISCUSSION

Based on this preliminary analysis, the results seem to suggest that the items related to perceived workload are relatively internally consistent and are valid. That is, the response characteristics for the items related to workload appeared to have a high degree of internal consistency and the relationships between these items correlated significantly in the direction expected. For example, BNUM Question 1 and SEQ were positively correlated ($r = 0.4$), meaning that as SEQ ratings increase (i.e., denoting less perceived difficulty), so did Question 1 from BNUM (i.e., also denoting less perceived demand). The inverse can be said for relevant items from NASA-TLX to Question 1 from BNUM, as expected.

On the other hand, the reliability and validity of perceived situation awareness seemed to be less notable. Cronbach alpha indicate weaker internal consistency (i.e., as a measure of reliability) and fewer items are seen to be significantly correlated when reviewing the teal set of indications in Figure 4. For the items that did correlate, did so in the direction that was expected. For instance, as SART items related to understanding of the situation increased, so did Question 2 from BNUM indicating greater perceived situation awareness.

The shortcomings observed in reliability and validity for situation awareness may be indicative of the challenges known within the human factors community in measuring situation awareness from a subjective questionnaire (e.g., Endsley et al., 1998). That is, it may be difficult to report perceived situation awareness on conditions that one is not aware of. Perhaps an alternative interpretation of SART is the degree of confidence one has in performing their tasks given their understanding of the situation and workload (i.e., perceived quality of situation awareness; Endsley et al., 1998). Additional data is needed to further understand the relationship of SART with BNUM Question 2; though, the preliminary findings presented in this paper show promising trends that may suggest that the BNUM survey provides valid and reliable indication of perceived workload and situation awareness.

CONCLUSION

This work presents an exploratory evaluation of the validity and reliability of the BNUM survey using real-world data collected from three separate and independent human factors engineering studies, utilizing licensed operators from operational nuclear power plants. While the results presented here are preliminary and should be interpreted with caution due to the limited sample size, the findings from this work appeared promising, particularly in BNUM's capacity of measuring perceived workload. To further validate BNUM future research will continue to collect additional data to support confirmatory factor analysis. It is the author's view that, when used as part of a multi-method and multi-measure approach, BNUM may offer a practical way of addressing human factors engineering considerations related to workload and situation awareness assessment in future large-scale digital modifications at existing nuclear power plants, as well as the design and development of advanced reactor technologies. BNUM offers a means of quantitative assessment of these two human factors constructs, as well as providing diagnostic criteria that can be used to inform potential shortcoming in the design of human-system interfaces, procedures, and training that ensures safe and reliable operation of such prospective nuclear power plants.

ACKNOWLEDGMENT

This work of authorship was prepared as an account of work sponsored by Idaho National Laboratory (under Contract DE-AC07-05ID14517), an agency of the U.S. Government. Neither the U.S. Government, nor any agency thereof, nor any of their employees makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.

REFERENCES

Braarud, P. Ø. (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.

Endsley, M. R., Selcon, S. J., Hardiman, T. D., & Croft, D. G. (1998, October). A comparative analysis of SAGAT and SART for evaluations of situation awareness. In Proceedings of the human factors and ergonomics society annual meeting (Vol. 42, No. 1, pp. 82–86). Sage CA: Los Angeles, CA: Sage Publications.

Kassambara, A. (2023). ggcorrplot: Visualization of a Correlation Matrix using 'ggplot2'_. R package version 0.1.4.1, <https://CRAN. R-project.org/package=ggcorrplot>

Kovesdi, C., & Joe, J. (2019, February). Exploring the Psychometrics of Common Post-Scenario Human Factors Questionnaires of Workload, Situation Awareness, and Perceived Difficulty. In *11th Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies, NPIC and HMIT 2019* (pp.1631–1643). American Nuclear Society.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In Advances in Psychology (Vol. 52, pp. 139–183). North-Holland. https://www.eia.gov/pressroom/releases/press542.php

Joe, J. C., & Kovesdi, C. (2021, June). Using qualified on-site nuclear power plant simulators in human factors validations of control room upgrades. In *International Conference on Applied Human Factors and Ergonomics* (pp. 171–178). Cham: Springer International Publishing.

Matthews, G., & Reinerman-Jones, L. (2017). *Workload assessment: How to diagnose workload issues and enhance performance*. Human Factors and Ergonomics Society.

Revelle, W. (2023). psych: Procedures for Psychological, Psychometric, and Personality Research_. Northwestern University, Evanston, Illinois. R package version 2.3.6, <https://CRAN. R-project.org/package=psych>.

Sauro, J., & Lewis, J. R. (2016). Quantifying the user experience: Practical statistics for user research. Morgan Kaufmann.

Taylor, R. M. (1990). Situational Awareness Rating Technique (SART): The development of a toll for aircrew systems design. In: AGARD Conference Proceedings No 478, Situational Awareness in Aerospace Operations. Aerospace Medical Panel Symposium, Copenhagen, 2nd -6th October 1989.

U. S. Nuclear Regulatory Commission. (2012). Human Factors Engineering Program Review Model, NUREG-0711, Rev. 3. Washington, DC: U. S. Nuclear Regulatory Commission.

U. S. Nuclear Regulatory Commission. (2015). Workload, Situation Awareness, and Teamwork. NUREG/CR-7190, U. S. Nuclear Regulatory Commission.