

EXPLORING THE USE OF COGNITIVE MODELS FOR NUCLEAR POWER PLANT HUMAN-SYSTEM INTERFACE EVALUATION

Casey R Kovesdi, Jeffrey C Joe

October 2019



The INL is a U.S. Department of Energy National Laboratory
operated by Battelle Energy Alliance

EXPLORING THE USE OF COGNITIVE MODELS FOR NUCLEAR POWER PLANT HUMAN-SYSTEM INTERFACE EVALUATION

Casey R Kovesdi, Jeffrey C Joe

October 2019

**Idaho National Laboratory
Idaho Falls, Idaho 83415**

<http://www.inl.gov>

**Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517**

EXPLORING THE USE OF COGNITIVE MODELS FOR NUCLEAR POWER PLANT HUMAN-SYSTEM INTERFACE EVALUATION

Casey R. Kovesdi, M.S. and Jeffrey C. Joe, M.S.
Idaho National Laboratory

Many of the United States' commercial fleet of nuclear power plants (NPPs) are approaching the end of their operating licenses. To extend the life of these plants, advanced human-system interface (HSI) technologies are being researched to address aging and reliability concerns with existing legacy systems. Human factors engineering (HFE) a critical role in ensuring these technologies are designed in a way that does not introduce new failure modes and promotes optimal human-system performance. An important focus of HFE in NPP modernization is early involvement to inform design of the HSI. This includes traditional formative evaluations, which are done to collect design feedback. While these evaluations are useful, they are limited in providing convincing quantitative data for efficiency of use. This paper discusses the use of cognitive models to provide quantitative data early in the HSI design process. A comparison is made of three open-source and readily accessible cognitive modeling tools.

INTRODUCTION

The existing United States (U.S.) commercial nuclear power plant (NPP) fleet generates about 20% of the nation's electricity. However, many NPPs within the U.S. fleet are now approaching the end of their operating licenses. The Department of Energy Light Water Sustainability Program (LWRS) is active in conducting targeted research and development (R&D) to extend the operating lives of these plants by ensuring they stay cost-competitive and maintain safe, reliable, and efficient operation. One LWRS pathway is researching digital instrumentation and control (I&C) technologies that address aging and reliability concerns with existing legacy I&C for these NPPs.

Advanced human-system interface (HSI) technologies are being researched to replace the legacy analog instrumentation and controls that once covered the entire footprint of the main control room. These advanced digital systems offer new capabilities with increased levels of automation for control actions, information and decision aiding, interface management, and administrative tasks (EPRI, 2015). The inclusion of human factors engineering (HFE) is crucial in ensuring these technologies are designed in a way that does not introduce new failure modes and that promotes optimal human-system performance (NUREG-0711, 2002).

As with other engineering domains, the HFE process requires advance specification of system characteristics followed by systematic demonstration that the as-built system has these characteristics (Good et al., 1986). For NPP modifications, the U.S. Nuclear Regulatory Commission follows this approach by systematically reviewing an applicant's HFE program as described in NUREG-0800 (Rev. 3), *Standard Review Plan Chapter 18 - Human Factors Engineering* (2016), and NUREG-0711 (Rev. 3), *Human Factors Engineering Program Review Model* (2012). The review criteria in NUREG-0711 suggests that applicants integrate HFE in the development process early and throughout the modification's lifespan, with an emphasis on later-stage verification and validation (V&V) efforts like integrated system validation (ISV).

Recent work, namely the Guideline for Operational Nuclear Usability and Knowledge Elicitation (GONUKE) framework, provides a detailed process that utilities can follow to ensure successful V&V. GONUKE emphasizes a cost-effective approach by focusing on early-staged HFE evaluations that are most impactful in influencing and refining design, as opposed to later-stage ISV of a final system (Boring, 2015). In its simplest form, GONUKE suggests *formative* and *summative* evaluations, each using methods in *verification* and *validation*.

While summative evaluation is synonymous with ISV as described in NUREG-0711, formative evaluations provide a basis for refining the HSI design through rapid iterative testing and evaluation. Common methods at this phase include focus groups, usability tests, surveys, and other forms of contextual research. These methods often rely on smaller sample sizes than those used in summative evaluation, as their statistical basis is fundamentally different (e.g., Nielsen, 1994). Plant personnel (e.g., licensed operators) who participate in these formative evaluations are not formally trained on the HSI as they would be for their licensing requirements. Hence, their interactions with the HSI may be more exploratory in nature to provide input into the design requirements (e.g., Ulrich, Boring, and Lew, 2018).

In some cases, the HSI designer might be interested in understanding how an HSI concept may support efficient operation. For instance, there may be circumstances in which quick access to soft controls on the HSI is critical (e.g., access to a reactor trip button); there may even be a specific benchmark or time requirement for accessing these controls. While the current formative methods might help the designer understand qualitative aspects of the design, these methods do not provide convincing quantitative data for efficiency of use. As such, additional easy-to-use evaluation tools that can uniquely inform HSI design early in this formative process ought to be explored to expand the HSI designer's evaluation toolkit. One promising methodology that may support provision of this early quantitative data entails the inclusion of applied cognitive models. The next section discusses a candidate cognitive modeling approach with three available tools under this umbrella.

CANDIDATE COGNITIVE MODELS FOR HSI DESIGN

A comprehensive discussion of all available cognitive models and their underlying theories is beyond the scope of this paper; however, one overarching methodology that has had a notable impact to HFE is GOMS (Goals, Operators, Methods, and Selection rules), developed by Card, Moran, and Newell (1980; 1983). GOMS provides a top-down approach to decomposing a task, starting at the user's goal and breaking the goal down into sub-goals. Goals and sub-goals are achieved by applying methods and selection rules. Using this GOMS framework, there are open-source and readily accessible tools for interface evaluation. These tools are described next.

Keystroke-Level Model

Traditional keystroke-level models (KLMs) provide a set of primitives to model a skilled user interacting with a computerized system in an error-free manner (Card, Moran, & Newell, 1980). Each primitive (e.g., an underlying task, such as moving a mouse cursor to a target) contains specific empirically-derived quantitative execution times. Examples of KLM primitives include: keystrokes (e.g., 80 milliseconds for best typist), pointing a target on a display using a mouse (1100 milliseconds), clicking a mouse button (200 milliseconds), homing the hands on the keyboard (400 milliseconds), and mentally preparing to execute an action (1350 milliseconds). These execution times can be summed to create an overall predicted task time.

The application of KLMs has traditionally focused on computer-based interactions; extensions of the KLM have been explored in different domains, such as the design of handheld devices, automotive information systems, cockpit design in aviation, and human-robot interaction (e.g., Luo & John, 2005; Pettitt, Burnett, and Stevens, 2007; Campbell, 2002; Drury, Scholtz, and Kieras, 2007). A straightforward case for using traditional KLM in HSI design is modeling the time it might take a trained operator to find the turbine trip soft control on any given HSI display. Figure 1 shows how a simple spreadsheet could be used to calculate a predicted task.

Goal: Click the 'Turbine Trip' button on the HSI display	
Operators	Time (milliseconds)
Mentally prepare action sequence.	1350
Move cursor to 'Turbine Trip' button.	1100
Click 'Turbine Trip' button.	200
Total Time	2650

Figure 1. Example KLM prediction for a simple task.

There are a couple of notable limitations to KLM. First, KLM assumes that the primitives are sequential and that there is a set order to completing the task. Hence, tasks that occur in parallel (e.g., monitoring an HSI display while completing a set of control actions) may be difficult to model. Further, when task success contains more than one path, the development of KLMs through selecting individual primitives

can be tedious. Finally, the KLM primitives may be too general to capture subtle differences between HSI concepts (e.g., the size of a button or location of an indication).

CogTool

CogTool is an open-source program that expands on the traditional KLM approach by using the ACT-R cognitive architecture to provide more precise predictions (John et al., 2004). For example, the generic KLM pointing primitive was replaced as a function of cursor distance and button size using Fitt's Law. CogTool also enables the HSI modeler to generate models using a graphical user interface to create storyboards of the task under evaluation (Bellamy, John, and Kogan, 2011). The training requirements are minimal. The HSI modeler can develop quantitative predictions by simply importing HSI design concepts (i.e., as images) into CogTool, adding CogTool widgets onto the imported images, and then demonstrating the task within CogTool as a storyboard.

Figure 2 shows the interface of CogTool. The top screenshot shows the imported HSI displays in CogTool with widgets overlaid on the key indications and controls under evaluation. The bottom screenshot shows the results of the storyboard (e.g., navigate to the “ELEC” screen), presented with a predicted task time and the underlying primitives.

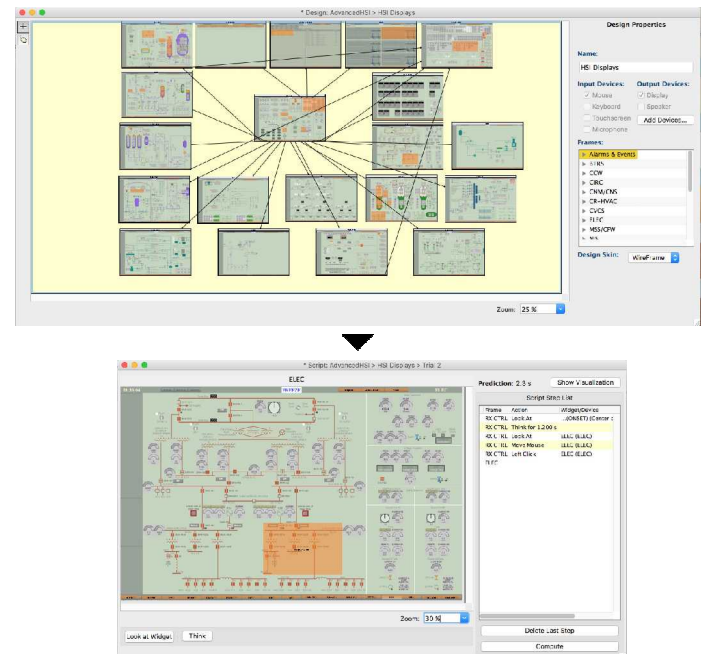


Figure 2. Example of the CogTool user interface.

CogTool was originally designed for computer-based tasks that involve viewing and interacting with information on a user interface and using a keyboard, mouse, or touch screen. As seen in the top portion of Figure 2 above, the available widgets are limited to mouse, keyboard, touchscreen, and microphone as inputs and a display and speaker for outputs. The modeler also has access to a timeline visualization of the generated ACT-R model in which there is an option to compare a second task or another HSI concept. The ACT-R

primitives shown in the visualization include system (frame), eyes (vision, eye movement preparation, and eye movement execution), cognition, and motor (see Figure 3).

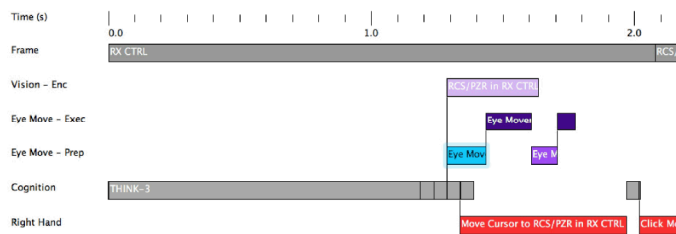


Figure 3. Example CogTool timeline visualization.

CogTool's predicted task times are claimed to be within 20 percent (i.e., +/- 10% of the point estimate) of observed human performance times (John et al., 2004). There is an option to present predicted performance in CogTool as a range to present potential variability in the prediction. Despite this accuracy claim, some literature has suggested that the accuracy of its predictions falls short for certain domains. For instance, Jorritsma and colleagues (2015) found that CogTool and other KLMs were not reliable in determining whether a given website design was faster than another.

Further, Adio (2014) compared observed task times to CogTool's predicted times for HSI displays designed for the petrochemical industry. Adio found that the ACT-R "think time," specified at a constant of 1.2 seconds, was too short of a duration for control room monitoring tasks. Adio's rationale for this discrepancy was that process control tasks often require plant operators to scan multiple monitors, retain information in working memory, and apply decision rules that are not built into CogTool's core modeling architecture. Another notable limitation of CogTool is that its interactions are constrained solely to computer applications. That is, interactions with certain physical controls, such as rotary dials and J-handles, are not built into the inputs available for model generation. Further, the time required for operators to walk board to board cannot be readily modeled in CogTool.

Cogulator

Cogulator is an open-source program that uses GOMS to generate predicted task times with extensions in predicting working memory load and mental workload (e.g., Estes, 2017; Stanley et al., 2017). Cogulator is script-based, requiring the HSI modeler to develop predictive models through its own syntax. Cogulator contains predefined primitives that are based on the original KLM framework, as well as other GOMS models: NGOMSL (Natural GOMS Language), Cognitive, Motor, Perceptual (CPM)-GOMS, Card, Moran, and Newell (CMN)-GOMS, as well as human-information processor (HIP). Each of these models have differing granularity for what level of detail each primitive represents. For instance, the primitives for traditional KLM are at a higher level (e.g., generic think time) than primitives for HIP (e.g., motor processor). Hence, the modeler must decide what level of detail is needed in the analysis.

It is worth noting that Cogulator's built-in primitives for various motor actions are more comprehensive than those within CogTool and traditional KLM. There are representative actions that would be expected in control room operations readily available in Cogulator. There are motor primitives for turning a dial or knob (e.g., rotary dial), as well as grasping an object with a hand (e.g., J-handle). One other advantage of Cogulator is its flexibility in allowing the modeler to create new domain-specific primitives as needed. Finally, Cogulator's NGOMSL framework supports modeling multitask activities.

Figure 4 presents the interface of Cogulator. To illustrate its capability of modeling multitask activities, Figure 4 presents a navigation task (e.g., go to the "ELEC" screen) along with a mental calculation task (e.g., determining the temperature change from an indication in the control room). The latter task requires mental arithmetic, which influences working memory load, illustrated as "2.4 chunks" in the top section of Figure 4. Cogulator also provides a timeline visualization of the primitives used along with a map of when working memory was most impacted in the task sequence.

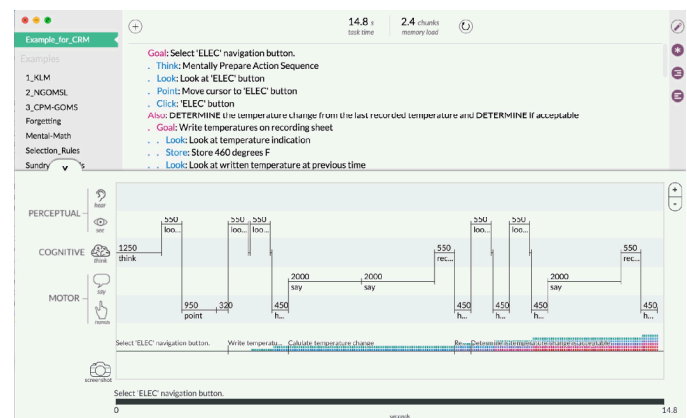


Figure 4. Example of the Cogulator scripting interface.

In general, Cogulator provides the most modeling flexibility, comprehensive modeling functionality, and available primitives; however, this inherent flexibility comes with the tradeoff of a steeper learning curve compared with KLM and CogTool. Cogulator may also be more tedious for creating models, as the HSI modeler must decide which GOMS model to use.

NOTABLE CHARACTERISTICS OF NPP OPERATION

The operation of NPPs has notable differences in interaction characteristics from the traditional commercial computer-based systems that initiated the use of GOMS in evaluation. For one, the main control room contains hundreds of indications and controls that span an entire room where control room operators must interface with multiple plant systems. Even in seated workstations, control room operators are required to interact with multiple monitors, each capable of presenting multiple displays. Second, control room monitoring tasks are often completed concurrently with other plant actions. For example, in a turbine startup, operators may

be required to complete a series of control actions while monitoring key parameters. Simple serial interactions with controls and indications would not fully capture the complexities of these concurrent interactions.

Third, control room operators are often required to store information into memory, compare and check the quality or validity of information from various indications, and perform mental calculations at certain points in a procedure. Simple think times may not accurately capture the cognitive intricacies of these tasks. Finally, control room operators work as a team, with each control room operator given a specific responsibility. Hence, information is often integrated among HSIs and exchanged among different crew members, as opposed to merely looking for information on a single information display.

COMPARISION OF COGNITIVE MODELS

Each cognitive modeling tool presents certain tradeoffs when applied to NPP HSI design. Table 1 below summarizes each of these tools' strengths and limitations from this context.

Table 1. Tradeoff evaluation of selected cognitive modeling tools.

KLM	CogTool	Cogulator
Strengths <ul style="list-style-type: none"> + Does not require specialized software + Little training required + Does not require a visual representation of the HSI. 	Strengths <ul style="list-style-type: none"> + Able to rapidly build models through storyboards (efficient) + Little training required with storyboard approach + Provides visualization of underlying cognitive processes + Offers ranges for predicted task times. 	Strengths <ul style="list-style-type: none"> + Does not require a visual representation of the HSI + Able to model concurrent tasks + Provides a more comprehensive list of primitives + Capable of creating domain-specific primitives if needed + Provides working memory load and workload predictions + Able to model cognitive tasks such as mental arithmetic and memorizing.
Limitations <ul style="list-style-type: none"> – Unable to model concurrent tasks – Assumes a specific order for activities in a task – Model creation can be tedious without running in formal software – Does not provide working memory or workload estimates – No built-in capability for visualizations to inspect the model – Model primitives may be too general to compare differences in design features across HSIs. 	Limitations <ul style="list-style-type: none"> – Requires a visual representation of the HSI – Unable to model concurrent tasks – Assumes a specific order for activities in a task – Provided primitives are specific to computer-based tasks – Difficult to add new domain-specific primitives – Does not provide working memory or workload predictions – No feature to model cognitive tasks such as mental arithmetic and memorizing. 	Limitations <ul style="list-style-type: none"> – Steeper learning curve – Model creation can be tedious using syntax – Model creation can be tedious due to syntax-based design.

POTENTIAL USES CASES IN NPP MODERNIZATION

Cogulator seems to provide the most comprehensive functionality for NPP HSI evaluation. Models in Cogulator are capable of considering multitask activities, workload estimates, and interactions with physical controls, as well as providing the option to add new primitives as needed. Cogulator may be applicable for modeling detection and monitoring tasks, situation assessment, planning responses, executing plant actions, and interface management tasks. This increased functionality comes at the cost of potentially increasing development time, which may threaten rapid design iterations. As such, care should be taken when scoping the use of Cogulator in the context of early HSI evaluation.

In contrast, CogTool offers an efficient means for building models through a point-and-click storyboard approach. Early wireframes and HSI mockups can be imported into the tool and rapidly modeled for predicted task times of basic interactions with HSI controls and indications. As such, CogTool may be helpful in part-task evaluation of specific interactions with an HSI concept. CogTool may be preferred for model interface management tasks due to its ease of use. Finally, the use of traditional KLM falls within the same application as CogTool (although CogTool offers a more sophisticated cognitive architecture and usable interface). Table 2 summarizes these potential use cases for the general KLM approach, CogTool, and Cogulator.

Table 2. Identified use cases for cognitive models in NPP modernization.

Cognitive Activities in a NPP	KLM	CogTool	Cogulator
Detecting and Monitoring	-	-	X
Situation Assessment	-	-	X
Planning Responses	-	-	X
Plant Actions (Control Room)	-	-	X
Plant Actions (HSI)	X	X	*Preferred
Interface Management	X	*Preferred	X

It must be emphasized that further research is needed on all three tools to better understand how they could fit within an iterative HSI design process. Particularly, future research should focus on modeling of tasks related to control actions that affect various plant systems. For instance, the level of granularity with primitives used in these tools must be better understood to the extent that they can provide meaningful data while also supporting rapid iterative evaluations. Recent work in human reliability analysis has studied the role of GOMS in dynamic human error prediction (Boring, Ulrich, and Rasmussen, 2018). Boring and colleagues distinguish between procedure-level primitives and lower-level task-level primitives. Within the context of HSI design, a procedure-level primitive may be quicker to model but may miss certain nuances of the HSI design. For example, modeling the interaction with a specific soft control may require a finer level of analysis.

Furthermore, additional validation of these tools is necessary to establish confidence in them as acceptable methods for NPP modernization. The collection of empirical data from future operator-in-the-loop studies of more detailed cognitive tasks, such as performing mental calculations,

decision-making, and concurrent monitoring, will inform future development of NPP-specific primitives that are necessary for modeling HSI interactions in a control room.

CONCLUSIONS

This work examined the use of KLM, CogTool, and Cogulator as potential tools for early HSI evaluations. Each of these tools provided unique and promising use cases. For instance, the use of KLM and especially CogTool may be favorable for quick evaluations of simple interactions with various controls and interface management tasks. Cogulator has greater functionality to support more comprehensive modeling of certain tasks such as when operators must make actions to control the NPP. HSI designers should consider these benefits and their corresponding tradeoffs to best support design questions on a case-by-case basis. To this end, future research is also needed to better understand the granularity, validity, and reliability of primitives used in GOMS models for NPP evaluation.

ACKNOWLEDGMENTS

This work of authorship was prepared as an account of work sponsored by Idaho National Laboratory, an agency of the United States Government. Neither the United States 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. Idaho National Laboratory is a multiprogram laboratory operated by Battelle Energy Alliance LLC, for the United States Department of Energy under Contract DE-AC07-05ID14517.

REFERENCES

- Adio, O.D. & Ikuma, L. Harvey, C., & Nahmens, I. (2014). Assessing cogtool time prediction accuracy on control room displays. In *IIE Annual Conference and Expo*, 3325-3334.
- Bellamy, R., John, B., & Kogan, S. (2011). Deploying CogTool: integrating quantitative usability assessment into real-world software development. In *Proceedings of the 33rd International Conference on Software Engineering*, 691-700.
- Boring, R. L., Ulrich, T. A., Joe, J. C., & Lew, R. T. (2015). Guideline for operational nuclear usability and knowledge elicitation (GONUKE). *Procedia Manufacturing*, 3, 1327-1334.
- Boring, R. L., Ulrich, T. A., & Rasmussen, M. (2018). Task level errors for human error prediction in GOMS-HRA. In *Safety and Reliability-Safe Societies in a Changing World*, 433-439.
- Campbell, C. B. (2002). Advanced integrated general aviation primary flight display user interface design, development and assessment. In *Digital Avionics Systems Conference*, 2, 10A5-10A5.
- Card, S. K., Moran, T. P., & Newell, A. (1980). The keystroke-level model for user performance time with interactive systems. *Communications of the ACM*, 23(7), 396-410.
- Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. CRC Press.
- Drury, J. L., Scholtz, J., & Kieras, D. (2007). Adapting GOMS to model human-robot interaction. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, 41-48.
- Electric Power Research Institute (2015). *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 for Operating Plants and New Builds*. EPRI, Palo Alto, CA. 3002004310.
- Estes, S. (2017). "Cogulator," The MITRE Corporation, [Online]. Available: <http://cogulator.io>.
- Good, M., Spine, T. M., Whiteside, J., & George, P. (1986). User-derived impact analysis as a tool for usability engineering. In *ACM SIGCHI Bulletin*, 17(4), 241-246.
- John, B.E., Prevas, K., Salvucci, D.D., & Koedinger, K. (2004). Predictive human performance modeling made easy. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 455-462.
- Jorritsma, W., Haga, P. J., Cnossen, F., Dierckx, R. A., Oudkerk, M., & van Ooijen, P. M. (2015). Predicting human performance differences on multiple interface alternatives: KLM, GOMS and CogTool are unreliable. *Procedia Manufacturing*, 3, 3725-3731.
- Luo, L., & John, B. E. (2005). Predicting task execution time on handheld devices using the keystroke-level model. In *CHI'05 extended abstracts on Human factors in computing systems*, 1605-1608.
- Nielsen, J. (1994). *Usability engineering*. Elsevier.
- Pettitt, M., Burnett, G., & Stevens, A. (2007). An extended keystroke level model (KLM) for predicting the visual demand of in-vehicle information systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1515-1524.
- Stanley, R. M., Kelley, D., Wilkins, S., & Castillo, A. (2017). Modeling the effects of new automation capabilities on air traffic control operations: Approved for public release; Distribution unlimited. Case number 17-2692. In *Digital Avionics Systems Conference (DASC)*, 1-7.
- Ulrich, T. A., Boring, R. L., & Lew, R. (2018). Qualitative or quantitative data for nuclear control room usability studies? A pragmatic approach to data collection and presentation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 61(1), 1674-1678.
- 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. (2016). *Standard Review Plan for the Review of Safety Analysis Reports for Nuclear Power Plants: LWR Edition—Human Factors Engineering, Chapter 18*, NUREG-0800, Rev. 3. Washington, DC: U.S. Nuclear Regulatory Commission.