Meta-Level Design Guidance and Operator Performance Measures for Hybrid Control Rooms

Reactor Concepts Research Development and Demonstration (RCRD&D)

Matthew Weinger
Vanderbilt University

Alison Hahn, Federal POC
Katya Le Blanc, Technical POC
Meta-Level Design Guidance and Operator Performance Measures for Hybrid Control Rooms

Final Report

September 2018

Center for Research and Innovation in Systems Safety
Vanderbilt University Medical Center
1211 21st Ave S, MAB Suite 723
Nashville, TN 37212

Sponsor:
Department of Energy

Grant Number:
DE-NE0008267
Sponsor: Department of Energy
Grant Number: DE-NE0008267
Project Title: *Meta-Level Design Guidance and Operator Performance Measures for Hybrid Control Rooms*
Principal Investigator: Matthew B. Weinger, MD, MS
Professor of Anesthesiology, Biomedical Informatics, and Civil & Environmental Engineering
matt.weinger@vanderbilt.edu, 615-936-6598
Submitting Official: D. Clinton Brown, Director
Office of Sponsored Programs
sponsoredprograms@vanderbilt.edu, 615-875-6070
Submission date: September, 2018
DUNS #: 004413456 (Medical Center)
Recipient Org: Vanderbilt University
3319 West End Avenue
Suite 100
Nashville, TN 37203
Project Period: 10/1/2014 – 9/30/2017
1. ABSTRACT ..................................................................................................................................................... 4
2. BACKGROUND .................................................................................................................................................. 5
3. ACCOMPLISHMENTS ....................................................................................................................................... 6

3. OBJECTIVES & RESULTS ................................................................................................................................. 8
   Gaining an understanding of nuclear power and operators work ........................................................................ 8
   Objective 1. Meta-level Design Guidance Document ................................................................. 8
   Results. .............................................................................................................................................................. 9
   Objective 3. Scenario Ontology Development and Verification ............................................................... 17
   Objectives 1-3 Evaluation. .......................................................................................................................... 22
   Objective 4. Data fusion model development and verification ..................................................................... 26

4. PUBLICATIONS & PRODUCTS ....................................................................................................................... 28

5. REFERENCES ..................................................................................................................................................... 30
1. ABSTRACT

The purpose of this grant was to refine and test methods to evaluate digital human-system interfaces (HSI) for existing largely analog nuclear power plant (NPP) control rooms (CRs). We hypothesized that CR operator performance (CROP) in hybrid NPP CRs will be improved if future digital HSI is designed based on: 1) more robust design guidelines for operator situation awareness (SA) and decision-making (DM); 2) enhanced CROP metrics during HSI evaluation; 3) an ontology for selecting simulation scenarios to evaluate digital HSI; and 4) a model to predict optimal HSI design attributes given CROP objectives.

The team conducted numerous nuclear power plant site visits (e.g., to Brown’s Ferry, Palo Verde, Sequoyah and Vogtle) to study control room operations and crisis management and 3 visits to the NRC Chattanooga Training Center. We observed research studies at DC Cook and NRC Chattanooga in collaboration with the Halden Reactor Project. We collected operator performance and workload data in one NPP simulator. A comprehensive human (i.e., CROP) performance literature review was published in Applied Ergonomics. We obtained operator team performance data from CAER to produce our data fusion model and a paper delineating the results have been accepted (revision submitted) in Resilience Engineering and System Safety. We analyzed interview data related to scenario ontology development; a manuscript is still in preparation.
2. BACKGROUND

**Project Background.** Vanderbilt University received grant DE-NE0008267 through the Department of Energy’s NEUP program in 2014. The project investigates the use of various human performance measures to evaluate advanced digital technology in the control room of nuclear power plants.

The nuclear power plant (NPP) operator is not only responsible for the health of the plant, but for maintaining critical safety functions that protect the facility, its workers and the public. The displays and controls within the NPP main control room are effectively the sensors and actuators of the operator enabling him to monitor, diagnose and manipulate plant conditions. Traditional analog human-system interfaces (HSIs) are distributed throughout the control room, usually based on subsystem and function. However, because of display salience, an experienced operator can often scan the entire control room and gain a holistic understanding (i.e., situation awareness or SA) of plant status.

As US utilities build new reactors and modernize plant infrastructure within their existing fleets, new opportunities and imperatives emerge for upgrading HSIs, increasingly using digital control-display technology. As a result, today NPP operators may encounter a purely analog control room, a fully digital control room or, most often, a hybrid of both digital and analog controls and displays. Regardless of the configuration, the HSIs of the NPP control room must support operator SA and decision-making for safely, effectively and efficiently managing plant performance at a tolerable cognitive load. Other complex, high-risk industries have also successfully begun to address this challenge.

Significant guidance exists to aid designers in NPP HSI development [e.g., NUREG-0700/r2(1) and EPRI Reports 1010042(2) & 1015089(3)]. However, there is limited formal consensus on what high-level (or meta) functions an HSI must support in terms of human performance in the NPP context [EPRI Report 1025791(4)]. The HSI must support operator SA and decision-making to assure safe, effective and efficient plant performance at an acceptable level of operator stress and workload, even under off-normal conditions. Thus, modernizing existing NPP control rooms present unique challenges to HSI designers, implementers, utilities, trainers, and control room operators.

An additional issue that must be addressed is the criteria for a “successful” HSI design. The critical criteria should be Control Room Operator Performance (CROP) during realistic simulations of normal, abnormal, and emergency NPP operations. However, what constitutes safe, effective, and efficient performance? With the implementation of new modular digital technology in a hybrid plant, there is a risk that operator performance, particularly on holistic SA and decision-making, can be affected adversely under some conditions if the HSI is inadequately designed and evaluated.

Like NPP operators, practitioners in other domains, including healthcare, aviation, disaster response, and the military perform in high-risk, high-consequence environments. All of these domains require workers to maintain an understanding of a dynamic and complex system, where the ability to diagnose and act quickly and accurately is critical.(5) Significant work has been undertaken in non-nuclear domains to measure and assess human performance, for example in healthcare,(6, 7) air traffic control, disaster response(8) and chemical process plants.(9) Substantial insight can be gained from an analysis of available literature in these domains, and used to inform HSI design for hybrid NPP control rooms. In
In this project, we are seeking to refine methods to evaluate how the introduction of digital HSI can best support the work of operators in existing hybrid NPP control rooms.

**Major Project Objectives.** This project was based on the premise that control room operator performance (CROP) in hybrid NPP control rooms will be improved if future digital HSI components are designed based on: 1) more robust design guidelines for enhancing operator SA and decision-making; 2) enhanced operator performance metrics during HSI evaluation; 3) simulation scenarios that are designed to most effectively evaluate specific digital HSI; and 4) a model that provides predictions of optimal HSI design attributes given operator performance level objectives. Thus, using a human factors engineering (HFE) approach, we are addressing the following project objectives:

1. Develop and validate meta-level design guidance (MDG) for NPP operator SA and decision-making relevant to the design of digital HSI components intended for hybrid control rooms. (In Year 1, we were asked by Idaho National Laboratory to focus on overview displays);
2. Identify, refine and validate additional human performance measures for SA and decision-making relevant to new digital HSI designs;
3. Create and evaluate a simulation Scenario Ontology for the selection of scenarios to evaluate new digital HSI designs; and
4. Create a data fusion model that supports heuristic evaluation of digital HSI designs whereby the design attributes are model inputs, the MDG and other design guidance are model operators, and performance measures are model outputs.

To achieve these Objectives, we integrated and applied knowledge and expertise from the healthcare industry, nuclear power industry, and other engineering disciplines (e.g., information sciences, civil, environmental) to begin to develop and validate new tools to evaluate digital HSI designs for hybrid control rooms.

### 3. ACCOMPLISHMENTS

Accomplishments throughout the three-year project include the following:

- Developed and fostered an effective multi-disciplinary research team (spanning 3 Departments at Vanderbilt (Anesthesiology, Computer Science, and Civil and Environmental Engineering), a collaborator from another academic institution (Nathan Lau from Virginia Tech), and researchers and domain experts at the Idaho National Laboratory (INL).
- Visited INL to meet with relevant personnel providing scientific oversight of the project (Hallbert, LeBlanc) as well as those conducting related research to foster collaboration and synergism with a second trip in year 2 to work with INL and Halden Reactor Project.
Through visits and correspondence, we strove to build collaborative relationships with the Human Factors Team of the Halden Reactor Project, NRC Chattanooga, Sequoyah NPP, Watts Bar NPP.

Conducted observations and observed simulations at Sequoyah NPP, Palo Verde NPP, Browns Ferry NPP, Vogtle NPP, DC Cook NPP and NRC Chattanooga. The notes from these observations were transcribed and coded.

Conducted interviews of simulation personnel at Sequoyah NPP, Palo Verde NPP, Vogtle NPP, and Watts Bar NPP. These interviews were transcribed and coded and have been used to advance the scenario ontology.

Used Bioharness™ systems to collect physiological data from NPP control room operators during simulation scenarios at Sequoyah NPP and at NRC Chattanooga.

Collaborated with Virginia Tech and the State of Virginia’s Center for Advanced Engineering and Research (CAER) center on the analysis of eye tracking and other operator and system performance data. These data were used to develop the hybrid data fusion model that we expect appear in a research article in Resilience Engineering and Systems Safety (currently under second round review after provisional acceptance).

Initiated potential collaborations with appropriate representatives at TVA, Dominion Power, and Palo Verde Nuclear Generating Station. None of these efforts bore fruit.

Provided input to study design of HSI testing for INL’s BENEFIT project with regard to the design of overview displays, computer-based procedures, and advanced alarms for hybrid NPP CRs.

Conferred with INL personnel about appropriate validation testing of new NPP control room HSI designs for incorporation into current national standards development efforts.

Presented three proceedings papers at the American Nuclear Society (ANS) 2017 Nuclear Plant Instrumentation and Control (NPIC) and Human Machine Interface Technologies (HMIT) Conference in San Francisco, CA – one on our literature review, one on meta-level design guidance, and one on the preliminary scenario ontology.

Presented relevant data on human error at the 2017 International Conference on Applied Human Factors and Ergonomics conference in Los Angeles, CA.

Created a software platform to support the conduct of a comprehensive literature review on operator performance measures in the process control industry. The resulting review will appear in the November 2018 issue of Applied Ergonomics.
3. OBJECTIVES & RESULTS

Gaining an understanding of nuclear power and operators work

The first six months of this project were spent organizing the team, arranging nuclear power plant site visits, planning and orchestrating the literature search, and drafting the human subjects protocol. The protocol to conduct interviews and observations at nuclear power plants was approved by Vanderbilt’s Institutional Review Board (IRB).

One member of the research team, Shilo Anders, traveled to Lynchburg, VA to take a 7-day nuclear power systems course offered by Areva to new employees to learn about nuclear power generation and components of the power plant. The research team subsequently traveled to Brown’s Ferry, Vogle, DC Cook, and Palo Verde NPPs to observe scenario exercises in the control room simulators and interview operators and simulation instructors who developed (primarily training) scenarios. In collaboration with Virginia Tech, the research team then traveled to Sequoyah NPP, Watts Bar NPP, and NRC Chattanooga to observe and collect data during scenario exercises, interview operators and interview scenario developers. We also toured Bellefonte NPP with a representative from the NRC.

Additionally, team members visited the NRC training center in Chattanooga three times to observe simulations on a number of different types of simulators. All of these activities allowed the research team to learn how hybrid nuclear power plant control room operators do their work and how HSI design may influence that work. These activities were essential to build our knowledge base while simultaneously achieving the grant Objectives.

Objective 1. Meta-level Design Guidance

Members of the research team visited INL, after which we consolidated the project aims to align better with the BENEFITS project (at Project Officer request). As a result, our meta-level design guidance was focused on overview displays with computer-based procedures a secondary objective. We therefore focused on the literature on large screen central overview displays. While these are widely used through many ‘process control’ and ‘command-and-control’ industries, the number of studies that have evaluated their use, much less their interface design, as a function of operational performance, was surprisingly small. We found almost no literature on the effects of distributed overview displays.

In parallel, we conducted interviews and observations with operators at three NPPs. The NPP operator participants ranged in experience from a few months post-licensure to >30 years of NPP operational experience. The older operators frequently were retired Naval submarine NPP operators, but the demographics of the younger operators varied from a recently retired Naval submarine NPP operator to those with scientific or engineering oriented undergraduate degrees. The majority of the observations occurred in the simulation center, as part of ongoing training with abnormal and emergency events. Observations were also conducted during normal plant operations. Interview participants consisted of five reactor operators (RO), one shift manager, and one shift technical advisor (STA). Control room personnel were opportunistically interviewed, while they were conducting their shift training or on other work duties.

Three researchers conducted observations and interviews during 3 different training scenarios – an emergency-plan scenario that included a hostile act, reactor startup, and steam generator tube rupture – ranging from 35 to 180 minutes duration were observed. The researchers observed
from either the simulator control room or within the simulated control room itself. Simulation instructors were available to answer researchers’ questions regarding specific interactions or events that occurred during the simulation. All researchers observed the pre-session trainer overview discussion. All notes were transcribed and coded.

Additional normal operations observations and interviews were conducted with the various control room personnel while they completed their normal shift work. After receiving individual consent, interviews were 15 to 60 minutes duration, and were interrupted by typical shift responsibilities. The interview guide focused on overview displays and computer-based procedures. Specific interests included evaluation of a system composed of distributed displays that was frequently used in a manner similar to overview displays and its associated functionality and limitations. Interview questions also queried personnel regarding their expectations for large screen overview displays and information requirements for that type of system. Finally, a series of questions elicited feedback regarding the potential content and best uses of computer-based procedures.

Two team members qualitatively coded the resulting data to extract design requirements. The two coders iteratively met to confer on a categorization scheme and to resolve categorization discrepancies. Both researchers coded all of the observation and interview transcripts.

Results

The qualitative results from both observations and interviews are decomposed into a number of relevant human factors considerations by overview displays and computer-based procedures. Each primary categorization is further decomposed into specific relevant observations and feedback that can impact system design.

Overview displays. The observed NPP control rooms do not specifically incorporate large overview displays; however, a reporting system that was available on existing computer screens was frequently repurposed and effectively functioned as an overview display. This system incorporated six digital displays distributed around the control room; thus, the following results view this system as a distributed overview display system.

The displays need to be viewable throughout the control room. Given the configuration of the distributed displays, some operators had certain information (e.g., trend charts) displayed on one monitor and other information displayed on a second monitor across the room. One operator was observed sitting in the center of the room gathering general trend information from the distributed displays. When detailed information was required, physically walking to the particular display that had the necessary information displayed.

The layout of the overview displays directly impacts personnel’s situational understanding, cognitive workload, and decision-making. Given that the current system displays are not designed specifically for use as system overview displays, control room personnel commented on a number of attributes they deem necessary for true overview displays. The information layout on the digital displays needs to match the physical system boards. Previous systems depending on the usage information were reported as having missing information from the system or the layout was reversed from the board layout. The existing system also limits what information can be displayed concurrently when compared to the information operators required to support their responsibilities. For example, the feedwater system is displayed as a mirror image of the actual physical control room board layout. Some limits are necessary to ensure that a consistent layout is provided, but the personnel wanted systems that also provide flexibility. It
was quite cumbersome with the existing system to locate the information that may be needed or desired within the layout.

It is important that the displays, whether centralized or distributed provide the necessary overview content. The data collection activities confirmed results from the existing literature that cites trend charts as an important element of overview displays (Burns, et al, 2008). A limitation of the observed existing system was that only a small number of trend charts, with few data points can be displayed simultaneously. Further, it is critical that overview displays, whether centralized or distributed provide the ability for personnel to understand the currently displayed information easily and “at-a-glance”, while completing other primary tasks. The STA regularly ‘walks the systems’ and needs to know what is going on quickly during that data collection activity; thus, even if the distributed displays are set to display different information, the STA must be able to quickly assess the system state across the screens. For example, one operator mentioned that at the start of the shift, he walks the boards not only to verify all information states but also to verify that the distributed displays contain the information he expects them to contain, particularly when trend charts containing different information are displayed on multiple displays. This verification is important for his ability to look across the control room from a centralized location and assess the current system state. All interview participants frequently mentioned this criterion for overview displays.

The existing system was difficult to navigate when seeking to access information. For example, the system required a complex series of steps to “drill down” into the subsystems to access specific information. Further, the amount of available information is vast, which can make recalling where desired information is located within the system difficult. As a result, a large number of steps and a long period of time can be required to access desired information.

A concern raised regarding the representation of the NPP system on overview displays was overall system complexity. This complexity had a direct impact on the system mimics provided on the existing boards, often resulting in abstractions that could hide complexity. Thus, a concern mentioned by several interviewees was that digital displays may still be unable to represent accurately and effectively the very complex subsystems.

The usage context dramatically impacts the frequency with which the displays are used for overview purposes. It was observed that during normal operations, the displays were primarily used to provide trend charts that were checked on a regular, but low frequency basis. However, the use of the displays to provide overview related information increased significantly during the abnormal and emergency training simulations. Thus, operators thought that procedure specific screens (e.g., startup) in an overview display would better support their work.

The existing system was developed in the early-1990s, a time when the computer processing and interface display capabilities were significantly limited compared to today, leading to limitations related to this system’s flexibility. The system provides a number of preset displays; however, operators found that these displays were not designed for working on a specific task; thus, necessitating opening multiple screens during the course of a single task. The current system allows operators to customize some of the presented data on some displays (e.g., trend charts), but does not support fully customized data display combinations. An operator cannot save a customized screen for later use; if the operator leaves a customized screen to complete another task, that screen is “forgotten” and must be recreated the next time the operator desires it. Further, one is unable to access historical data that ARE no longer displayed on a particular screen, such as information that has moved off the end of a displayed trend chart. Note that
historical data are available via other systems. All of these aspects can directly impact decision-making.

The existing system also failed to provide sufficient decision support capabilities. Decision support systems typically integrate current and historical information that can be used by the system for predictive modeling, a critical aspect for making decisions that ensure NPP system safety. The plant had an industrial standard historical data management system that permits access to historical data points, but it was not integrated with the digital display system. Thus, personnel must integrate information across multiple systems and calculate, either by hand or by self-developed spreadsheets, predictions for certain procedures. The existing system does not provide personnel with tools that allow them to combine the current system state with predictive models so as to understand better the positive or negative potential outcomes of different system manipulations.

The inclusion of new information sources and communication media was deemed important when developing next generation overview displays. An important concern for many operators was the ability to visually check on areas in the physical plant to ensure it was safe to proceed with particular activities and to verify the safety of distributed personnel. Thus, the inclusion of remote controllable camera feeds throughout the plant were deemed important to achieve this objective. Further, instantaneous communication media, such as WIFI-enabled texting, were considered valuable capabilities.

As with any intelligent system, there is a concern about personnel becoming over reliant on overview displays, which could result in skill degradation. This concern has been demonstrated to be legitimate with the existing systems. It was reported that during prior training sessions, when certain non-safety critical systems were disabled, personnel failed to successfully complete the emergency recovery procedure. That said, during one of the observed scenarios which contained such a system failure, CR personnel were able to obtain the necessary information to complete the scenario successfully without the availability of non-safety critical system that they frequently used as if it was an overview display.

Centralized, large wall-based overview displays can provide additional information, but appear to have their limitations. An attractive alternative is the use of larger distributed displays specifically designed as overview displays that provide the information operators feel is missing from their current makeshift overview display system. Trend charts have been noted as an important element of large overview displays; Operators specifically requested more trend charts and the ability to have more flexibility in manipulating and customizing the data contained in the trend charts. Further, operators wanted information to be more readily accessible.

Computer-Based Procedures. The current procedures for all types of actions (e.g., regularly scheduled maintenance to emergency procedures) performed in the NPP control room are delineated in paper-based procedures contained in numerous binders located in and near the CR. For example, for scheduled maintenance, the procedure is printed out and distributed. Then, a CR operator, usually in collaboration with field personnel, works through the paper procedure to complete the tasks. This approach may require the field operator to return to the control room to complete the procedure, Computer-based procedures (CBP), while used in many other industrial plants are not yet prevalent in NPPs. We conducted interviews to generate data on design themes operators consider necessary for successful integration of CBP.
An important aspect of CPB functionality is flexibility. This speaks directly to developing something beyond a flat file or even a slightly enriched format (e.g., PDF forms). One operator described trying to complete a routine procedure via a PDF file using his tablet. Due to the inflexibility of the flat-file-based procedure, he quit and went back to paper. Current paper procedures allow operators to record values, make notes, and document completion of steps on a dated procedure document. Operators were acutely aware that during emergency procedures, the plant state will likely require deviation from the written procedure (e.g., beyond design basis) and that this must be documented. Other important requirements for flexibility include place keeping and documenting skipped steps accurately.

Like overview displays, the CPB’s content was mentioned frequently as a concern. Specifically, operators viewed the CPB’s as an opportunity for providing real-time system information directly associated with the procedure steps, which was viewed as a means to improve their efficiency and accuracy when completing a procedure. This information may include actual system component values or operator calculations, currently done by hand or by using other computer systems. Further, alarm-specific (or triggered) CPB can include set point information and allow the operators to see the parameter ranges and the steps required to address the alarm. The ability to integrate data from multiple sources and display it collectively in a context relevant space on the CBPs was deemed of high value to operators. These aspects all have the potential to contribute to better operator decision-making.

Additionally, operators felt that CPB could be used to navigate more efficiently alarm procedures that occur during operations. The operators expressed interest in using technology to help them complete the alarm procedures and even use smart alarms directly connected to the CPB to both bring up applicable procedures and to help prioritize them.

Operators discussed the new technology options which could support CBPs. Tablet-based CBPs were viewed as a viable and preferred means to accessing and completing procedures. Operators felt current alarm procedure binders should be replaced with a tablet. Currently, alarm procedures binder for a particular board are currently housed on the front of that board. Each binder could be replaced with a tablet that would be located in the same place and dedicated to that board. Further, operators thought a tablet-based version of surveillance or routine maintenance procedures had the potential to improve efficiency by incorporating real-time information from the field, via text or video feed to support whether and when to proceed with the next steps.

Summary. This qualitative evaluation employed standard human factors observation and interviewing techniques to understand how NPP personnel interact and use the technology in the control room to accomplish their work, with a specific focus on overview displays and computer-based procedures. Additionally, this evaluation elicited requirements for future control room system designs. Overall, the implementation and use of newer technologies appealed to operators. The current standard for evaluating technologies and training operators on those technologies occurs in the simulated control room. The majority of the time operators spend in the simulator is focused on responding to abnormal or emergency procedures, while the bulk of the operators’ work in the actual control room is conducting routine operations that create safe power generation. Designing and evaluating technology based on these highly infrequent extreme events may create technologies that irrelevant. Further, these new technologies may be burdensome to operators during normal operations, which can potentially increasing workload, hinder decision-making and negatively impact situation awareness.
Objective 2. Nuclear Power Plant Control Room Operator Performance (CROP) Measures

We reviewed the available literature on measuring human (operator) performance to assess the design of HSIs, focused on NPP control rooms, so as to identify best practices and future directions for research and operational improvement. We conducted a comprehensive review of all available studies related to human performance in the context of process control system design and evaluation.

This review focused on the power generation and petrochemical processing industries and included two types of studies: Evaluation Studies and Methodology Studies. Evaluation Studies evaluate one or more HSI components, such as overview displays, computer-based procedures (CBPs), alarms, control interfaces, or decision aids. The review of Evaluation Studies provided an overview of what and how human performance was measured. Methodology Studies focused on the development and evaluation of human performance measures or methods for HSI evaluation. They provided information regarding the qualities of the measurement instruments.

Inclusion and exclusion criteria. An article was included if it: 1) described one or more empirical studies of professional control room operators in the power generation and petrochemical processing industries; 2) described an Evaluation or Methodology Study; 3) was conducted in a high-fidelity simulation environment; and 4) included quantitative or qualitative human performance measurement data. An article was excluded if it: 1) was not written in English or 2) only described a research protocol without any results.

Search strategy. A library information specialist and two human factors experts developed the search queries. The search was a three-stage approach, starting with electronic databases, followed by the grey literature sources and expert inputs, and concluded with references listed in any of the included articles and those citing any of the included articles. Only publications between 1986 and 2016 were included.

The search combined terms from human performance and research domains. Human performance terms included: human performance, operator performance, crew performance, workload, situation awareness, decision-making, and human error. Research domain terms included: process control, nuclear power plant, chemical processing, petrochemical, healthcare, aviation, and transportation (articles not relevant to the process control domain were subsequently excluded from the analysis described in this paper). Terms in the same area were combined with the “OR” operator, while terms in different areas were combined with the “AND” operator.

The search strategy and results are delineated in Figure 1. Electronic database searched were IEEExplore, Ei Compendex, PsycINFO, Web of Science, and PubMed. The grey literature
search was conducted in the International Nuclear Information System (INIS) repository. Team members and colleagues also provided relevant reports and documents, for example, from the United States Nuclear Regulatory Commission (NRC) and Institute of Nuclear Power Operations (INPO). After all the identified articles were screened, reviewed, and coded, Google Scholar was used to identify additional cited and citing articles.

![Flow diagram of the study identification, screening, review, coding, and synthesis process of the current systematic review.](image)

**Figure 1.** Flow diagram of the study identification, screening, review, coding, and synthesis process of the current systematic review.

We systematically reviewed the literature on the human performance measures used to evaluate HSI in high-fidelity simulations in the process control domain. The most notable findings are: 1) The majority of the studies were conducted in dedicated research simulators in the nuclear power domain; 2) Most studies had small sample sizes (20 or less); 3) The human performance measures addressed six human performance dimensions (i.e., task performance, workload, situation awareness, teamwork/collaboration, plant performance, and other cognitive performance indicators); 4) The most frequently measured human performance dimensions were task performance, workload, and situation awareness; 5) Almost three quarters of the Evaluation
Studies measured human performance across multiple dimensions; 6) Some studies used multiple instruments to measure one human performance construct; and 7) Only a few measures, all of which were developed for this domain, showed evidence of acceptable reliability, validity, and sensitivity. At least in the process control domain, the available literature has gaps that preclude a full understanding of the measurement properties of the commonly used measurement instruments or provide a basis for selection of the most appropriate instruments for a particular study’s objectives and design.

Measuring the full spectrum of human performance

Traditional human factors research studies typically measure performance with “the big three” (10): speed, accuracy, and attentional demand. Indeed, our review of studies in the process control domain found that operational efficiency, operational accuracy, and workload were frequently used measures of human performance. However, additional human performance dimensions and constructs have emerged with advances in HFE theory and practice. From a systems perspective, human performance should be understood as the mediator of system inputs and outputs. Thus, constructs such as situation awareness and teamwork are important dimensions of human performance that should be measured in HSI evaluation studies. Consistent with this framework, several dimensions of human performance are considered critical for human factors validation of control room HSI (11, 12). The six human performance dimensions that emerged from our review provide a more comprehensive view of the transformation process.

In practice, however, only a small minority of studies measured human performance indicators across more than three performance dimensions. It is possible that, at least in some studies, more dimensions were measured, but were not reported. The use of a comprehensive battery of human performance indicators can be challenging, especially under time and budget constraints. Some measures require the collection, analysis, and interpretation of large volumes of data (e.g., voice communication, eye tracking and other physiological parameters, video analysis), which may not be feasible given study objectives and constraints. Measurement instruments that require effort by subject matter experts (SMEs) are particularly time and resource intensive. For example, OPAS requires extensive SMEs involvement in scenario analysis and for real-time performance scoring (13). Communication content and pattern analysis requires both domain and methodological expertise (14).

Another impediment to measuring performance across multiple dimensions is participant response burden. Response burden can be decreased, for example, with the use of more abbreviated instruments, intermittent or staggered sampling, and less frequent sampling. However, these approaches have limitations that become most problematic with the smaller sample sizes found in the existing literature. One can shift the burden from the participant to the researchers, for example, by measuring behavior with observational instruments and to measure workload using physiological indicators. As one example of staggered sampling, freeze-probe techniques to measure situation awareness (15, 16) instead of retrospective self-report ratings (17) can reduce response burden at the end of a study scenario when other self-report measures must be administered. Another consideration is measurement intrusiveness. Eye tracking or brain activity measurement, for example, can require participants to wear uncomfortable head gear. Such measures are less well tolerated, especially during long studies.

Measurement instruments and their measurement qualities for human performance

Several human performance measures were developed specifically for use in high-fidelity simulation studies in the process control domain. Examples of such measures include OPAS for
overall performance (13), COgnitive–Communicative–Operative Activity (COCOA) for workload (18), Situation Awareness Control Room Inventory (SACRI) and Process Overview for SA (15, 19-21), and Halden Cooperation Scale (HCS) for teamwork/collaboration (22). While limiting comparison of results across domains, there are several advantages to using domain-specific instruments in HSI evaluations. First, process control operators’ work in realistic environments is a relatively unique set of related tasks that are performed under domain-defining critical conditions (23). A well-designed domain-specific instrument can provide a valid, reliable, and sensitive measure of context-dependent changes in human performance. Indeed, evidence of reliability, validity, and sensitivity were available for many of these measures when used in high-fidelity simulations. Second, domain-specific instruments are more likely to yield meaningful diagnostic information to inform design improvements. This property is particularly useful in formative evaluations where the focus is not only identifying problems, but also diagnosing the nature of these problems and developing design solutions (24). For example, a domain-specific SA instrument, such as Process Overview (20), can help to ascertain if an operator can access, build, update, and project specific contextual information (e.g., a critical plant parameter) identified by process experts. If operators consistently fail to detect the critical situation during formative simulation studies, this would provide highly specific guidance for the redesign of the associated HSI elements.

Both NASA TLX and SART are widely accepted, domain-independent instruments (17, 25) that are low-cost and easy to administer (26). However, the frequency of usage does not necessarily reflect the quality of a measure (27). In the reviewed Methodology Studies, NASA TLX showed weak sensitivity and SART showed weak internal consistency and concurrent validity. Even if these common instruments had well-demonstrated excellent measurement properties in other domains, because their reliability and validity can be highly context-dependent (28, 29), more work is needed to establish their applicability and usefulness in high-fidelity simulations in the process control domain. In the literature, NASA TLX is reported to have high concurrent validity with behavioral performance (30); however, a number of studies showed that it was less sensitive than other instruments (e.g., Workload Profile and Subjective Workload Assessment Technique (SWAT)), particularly when comparing workload under different conditions, such as task difficulty (30), context effects (31), and interface output modality (32). Thus, when a study’s goal is to compare the effect of alternative designs on workload, rather than to predict behavioral performance, NASA TLX may not be the best choice. Across several studies, SART correlated well with participants’ perceived task performance and confidence level in their own SA, but not their actual SA (17, 26, 33). Furthermore, SART does not consider specific elements or parameters of the task environment. Thus, it may only be a good choice in environments where SA-related elements are difficult to define (33). In contrast, in process control rooms, one can readily define SA-related elements for given scenarios (16, 34). Here, it may be more appropriate to measure SA using a domain-specific freeze probe technique, especially given the often relatively slower (minutes to hours) pace of process control critical events, compared with other domains, such as aviation, aerospace and anesthesiology.

The usefulness of any measurement instrument is limited by its measurement quality. This may be particularly important when study sample sizes are limited. In this review, the instruments’ measurement properties were typically inadequately evaluated, not reported or, when reported, they were often below widely accepted criteria. Furthermore, some relevant measurement properties, such as specificity and measurement error, were not assessed in any of the reviewed studies.
Work Products


Objective 3. Scenario Ontology Development and Verification

The goal of this objective was to develop an ontology of NPP simulation scenarios to guide the selection of optimal scenarios for evaluating potential digital HSI modules for deployment in hybrid NPPs. ‘Scenario ontology’ means a formal categorical description of the concepts, attributes, goals, relationships, and conditions of NPP operation that are embodied in simulation scenarios and can influence NPP operators’ performance when using different HSIs.

A systematic literature review, focused on the process of scenario design and development, revealed 22 relevant articles, although these were largely in the healthcare domain. We developed an initial framework about how scenarios are developed. To refine and expand this framework, the research team interviewed simulation scenario developers at four NPPs and NRC Chattanooga. We used the interview guide to conduct semi-structured interviews with ten SMEs. Example questions included were: “What is your typical process for developing a scenario for training and/or evaluation?” and “What different types of scenarios exist? What makes them different?” We also observed eight simulation scenarios at four different NPPs. Information from NUREG-0711 (35) and feedback from the first round of interviews were used to supplement the interview guide in preparation for the next round of interviews.

In addition, we conducted a card sorting exercise with developers from two NPPs and from the Halden Reactor Project. The card sorting technique was developed to facilitate the creation and iterative verification of our scenario ontology. Through the analysis of the interview transcripts and card sorts, we further refined and expanded the ontology; We used more recent interviews to validate the scenario ontology framework that had been developed from the initial card sorts and interviews. Since almost all NPP control room simulations are used for CRO training and performance assessment, the ontology needed further refinement to achieve its aim of informing the design of simulations to evaluate NPP control room HSIs. Thus, we interviewed a SME from INL as final verification of the iteratively developed scenario ontology.

The interview data were collated and coded using grounded theory methodology (36-38). Three researchers used an iterative process of interview transcript review and coding, discussions to reach consensus, recoding, and refinement of the interview guide to solicit clarifying information during subsequent interviews. The researchers independently reviewed and coded each interview transcript, tagging all topics of potential relevance. During group discussions, they identified and reached consensus on individual scenario development process attributes, steps, goals, contingencies and measures, as well as hierarchical groupings, potential ordering, and all possible inter-dependencies. Uncertainties and disagreements were reformulated as questions and added to the interview guide. The hierarchical groups of attributes and goals were
then mapped onto the related procedural steps of the scenario development as a flowchart to capture similarities, discrepancies, and missing information for follow up studies. **Table 4** presents the attributes for consideration during the scenario development process for HSI evaluation.

**Results.**

**Figure 2** presents the first iteration of our scenario development process flowchart. The first step of creating a scenario is to determine the objective of the scenario. In this first iteration, three possible purposes (i.e., why) for scenario development were considered: CRO training, HSI evaluation, and system performance evaluation. Since we were primarily interested in HSI evaluation, much of the remaining ontology development focused on this topic. Because different types of HSI differ in terms of their functional requirements, operator demands and potential operational benefits and weaknesses, the next step is to specify which type of HSI is being evaluated. This drives consideration, at a high level, of the types of CRO tasks and cognitive functions to be engaged and evaluated. The next step is to determine how performance on these tasks and functions will be evaluated. Performance metrics, typically targeting operator and system performance as surrogates, are the subject of a separate paper. Nonetheless, the goal here is to determine how one will know that the HSI design is acceptable and, if flaws are detected, to ascertain sufficiently the nature of those vulnerabilities so as to mitigate them.

The next step (2a) is to identify the specific task requirements that will be induced by the HSI. For example, for CBPs, operators must look at a computer screen, identify their ‘place’ in the procedural sequence, determine the data acquisition or control tasks expected, and indicate (i.e., click or type) the requested data or check-off of completion. Here, the scenario developer wants to consider the interactions between operator and HSI that are most likely to lead to safety critical task failures.

The next step (2b) is to identify the simulated event characteristics or attributes that will best support the scenario objectives while considering the tasks to be performed and measures to be assessed. **Table 1** provides a list of event attributes discerned from the literature and verified during our interviews and observations. At this step, the developer is considering different scenarios that will contain the essential event attributes while engaging the CROs in the desired tasks with the HSI. Most scenarios are variations of secondary malfunctions that lead to major accidents, some of which are specified by the NRC (39).

**Table 1. Twelve Scenario Attributes Discerned from the Literature Review**

<table>
<thead>
<tr>
<th>Initial setup</th>
<th>Timing (e.g., expected duration)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial plant conditions</td>
<td>Complexity (e.g., no. of procedures used)</td>
</tr>
<tr>
<td>Perturbation (event)</td>
<td>Credibility (e.g., frequency of event)</td>
</tr>
<tr>
<td>Cognitive requirements for success</td>
<td>Severity (e.g., event type)</td>
</tr>
<tr>
<td>Practical distractions / interruptions</td>
<td>HSI Comprehensiveness (e.g., single system)</td>
</tr>
<tr>
<td>Herrings (red, blue)</td>
<td>Risk significant human actions</td>
</tr>
</tbody>
</table>

The next step (3) is to begin to design the scenario, codifying the event so as to meet the requirements defined in Step 2. At this ‘what’ stage, the scenario developer iteratively specifies an increasingly level of detail of the scenario’s evolution, considering the requirements and other...
contingencies of the prior stages. For this stage, our initial literature review identified five commonly used major transients and provides aspects of four event attributes – timing, complexity, credibility, and fidelity (Table 2).

Figure 2. Initial Scenario Development Flowchart
Table 2. Detailed List of Major Transients from Initial Literature Review

<table>
<thead>
<tr>
<th>Accident Description</th>
<th>Terrorist Attack</th>
<th>SGTR</th>
<th>Loss of Offsite Power</th>
<th>Loss of Main Feedwater</th>
<th>Turbine Trip</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timing</td>
<td>120 min</td>
<td>45 min</td>
<td>50 min</td>
<td>75 min</td>
<td>20 min</td>
</tr>
<tr>
<td>Complexity</td>
<td>High</td>
<td>High</td>
<td>Med</td>
<td>Med</td>
<td>Med</td>
</tr>
<tr>
<td>Credibility</td>
<td>Low</td>
<td>Med</td>
<td>Med</td>
<td>Med</td>
<td>High</td>
</tr>
<tr>
<td>Fidelity</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Med</td>
</tr>
</tbody>
</table>

Analysis of the interview transcripts provided a compendium of approximately 100 scenario attributes, which were divided according to their relevance to HSI evaluation. For HSI evaluation, the scenario attributes were organized under related steps of the scenario development process to augment the details discerned from the literature review (Table 1). Table 3 provides the scenario attributes extracted from the interviews, card sorts, complemented by our simulation observations.

Table 3. Scenario Attributes by Category

<table>
<thead>
<tr>
<th>Results from Literature Review</th>
<th>Detailed Attributes Derived from SME Interviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial setup</td>
<td>Precursors to planned event, initial conditions and associated attributes</td>
</tr>
<tr>
<td>Initial plant conditions</td>
<td>Initial plant conditions (typically but not always normal operations)</td>
</tr>
<tr>
<td>Perturbation (event)</td>
<td>Initial trigger such as an instrument malfunction</td>
</tr>
<tr>
<td>CRO cognitive requirements for success</td>
<td>Mitigation strategy</td>
</tr>
<tr>
<td></td>
<td>Situation awareness</td>
</tr>
<tr>
<td></td>
<td>Board awareness</td>
</tr>
<tr>
<td></td>
<td>Systems thinking</td>
</tr>
<tr>
<td>Practical distractions / interruptions</td>
<td>Other events (usually minor or unrelated)</td>
</tr>
<tr>
<td></td>
<td>Extraneous personnel</td>
</tr>
<tr>
<td>Herrings (red, blue)</td>
<td>Absence/wrong indication of control element</td>
</tr>
<tr>
<td>Timing (e.g., expected duration)</td>
<td>Scenario duration</td>
</tr>
<tr>
<td>Complexity (e.g., no. of procedures used)</td>
<td>Procedure selection</td>
</tr>
<tr>
<td></td>
<td>Multiple events</td>
</tr>
<tr>
<td></td>
<td>Number of events per scenario</td>
</tr>
<tr>
<td>Event timing (sequential vs. concurrent)</td>
<td>Unfamiliar/rare procedures or actions</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>Removal of key display system</td>
<td>Available mitigations</td>
</tr>
<tr>
<td>Credibility (e.g., frequency of event)</td>
<td>Based on real events</td>
</tr>
<tr>
<td>Frequent failure modes</td>
<td>Logical flow</td>
</tr>
<tr>
<td>Realism</td>
<td></td>
</tr>
<tr>
<td>Severity (e.g., event type)</td>
<td>Event evolution/rate of change</td>
</tr>
<tr>
<td>Magnitude of event</td>
<td>Personal risk (radiation, etc.)</td>
</tr>
<tr>
<td></td>
<td>Failed automation</td>
</tr>
<tr>
<td>HSI Comprehensiveness (e.g., single system, multi-system)</td>
<td>Intentionally left blank (see Discussion)</td>
</tr>
<tr>
<td>Risk significant human actions</td>
<td>Intentionally left blank (see Discussion)</td>
</tr>
</tbody>
</table>

The next step (4) is to have the scenario reviewed by SMEs and relevant stakeholders (e.g., NPP management, HSI vendor). In many NPPs, there are multiple levels of review and approval that must be integrated into the scenario development process, potentially at earlier or later Steps. The subsequently refined scenario would then be ‘finalized’ (Step 5) for pilot testing (Step 6) by mapping out the scenario flow and fleshing out all of the detailed simulator and simulation conditions necessary to run the scenario. Step 5 requires mapping out the scenario’s anticipated series of operator and plant responses, which include contingencies, distractions, and a success path. This “What if” step is critical to designing a scenario that will yield consistent results permitting aggregation of performance across operator teams. SME input, simulation instructor experience and brainstorming can be useful at this step to assure that the scenario addresses all conceivable operator responses to ongoing simulated events. Pilot testing (6) is absolutely crucial to validate technical assumptions (i.e., how the scenario will run) and expected operator behavior (e.g., unanticipated actions or inactions at various stages of the scenario), identify weaknesses or missing elements that degrade scenario reliability, and to assess whether the intended performance measures are likely to yield meaningful evaluation data for the targeted HSI (e.g., acceptable levels of sensitivity and specificity). It can be useful to present the results of these pilot tests to SMEs to gain insight into identified scenario shortcomings and unanticipated operator behaviors. As shown in Figure 1, the iterative nature of scenario development mandates feedback loops at various stages to further refine, and possibly change substantially, various aspects of the design.

We aimed to discover and verify the critical steps of a scenario development ontology using a flowchart with mapping codes, i.e., scenario attributes, extracted from the interview data for relevant steps of the process. The data extracted from the interviews appeared to support the Scenario Development Process delineated in Figure 1. However, future research will be necessary to further validate the sequence of steps within the flowchart, to identify missing steps, and to assure that the process is sufficiently generalizable for many types of HSI as well as for different NPPs. While we framed the interviews to elicit information about scenario creation for HSI evaluation, this may have limited or biased our results. Finally, the interviewees were inconsistently familiar (and less commonly operationally experienced) with the full spectrum of
modern HSI technologies under development. For example, the identified differences in the design process for scenarios for HSI design versus CRO training/assessment may actually be greater than observed.

**Work Products**


[NOTE: A journal manuscript is in preparation]


**Objectives 1-3 Evaluation**

In our original proposal, Objectives 1-3 had a combined evaluation that was proposed to be conducted at INL as well as in a NPP CR simulator. However, due to significant impediments to access to NPPs and especially to CR operators as study participants, we were unable to conduct the type and number of studies we originally planned. This was in part due to ongoing changes to INL’s Benefits project and their evolving relationship with Palo Verde NGS. Further, due to technical challenges with the INL simulator, the overview display study was postponed indefinitely. Neither INL nor our original industry partner were able to gain us sufficient access to NPPs and their operators to conduct the proposed studies. Nonetheless, we were able to collect initial data from 4 operators during 2 studies, once at the NRC Chattanooga and once at Sequoyah NPP (TVA). Despite initial assurances, we were never able to collect data at INL or Palo Verde.

Our team participated in the development of the scenarios to be used in the simulation studies that were to be conducted at INL in August 2016. The research team collaborated with the INL Benefits team and provided feedback on the overview display study that was to be Study 1. Our research team also visited INL on several occasions to assist in their CR modernization research efforts. Specifically, we worked with INL to develop scenarios that would be used in the aforementioned planned study to evaluate performance with and without an overview display. Independently (as a backup measure), we initiated a collaboration with the Human Factors Team of the Halden Reactor Project (HRP), participating in their studies at DC Cook and NRC Chattanooga. Due to circumstances beyond our control, the proposed study at INL was cancelled and we were not allowed to participate in the substitute workshop due to contractual constraints established by the participating NPP. By Year 3, it became evident that we would not be able to conduct any simulator studies at INL.

We were able to partner with HRP and the NRC Chattanooga for one study, and subsequently with Sequoyah for another to collect data from a total of seven operators in two different simulations. We had hoped to conduct more studies during the no-cost extension but have been unable to gain access to any NPP control room operators or convince NPPs to run simulations suitable for validation studies.
NRC and HRP Computer-Based Procedures Study.

The HRP study explored the use of computer-based procedures during a simulated emergent event at NRC. We obtained physiologic data on three participants as they participated in the scenario using a BioHarness (Version 1), a physiologic device that captures heart rate, breathing rate, skin temperature, body motion at a sample rate of 250Hz. Participants wore the device against their skin attached by an elastic strap around the chest. Prior to the start of the experiment, the BioHarness was attached to the participant and its function visually verified using the live-view function (via wi-fi to a research laptop). Then, participants were instructed to sit down and relax at their workstation with their eyes open to collect five minutes of baseline data. The simulation experiment proceeded with the device logging the physiological data locally. Researchers made notes throughout the simulation including when participants were particularly active and when the total loss of the CPS occurred.

Results.

The operator responses to the five test conditions in the scenario are summarized in Table 4.

Table 4: Summary of Operator Responses to Test Conditions

<table>
<thead>
<tr>
<th>Test condition</th>
<th>Failure description</th>
<th>Operator response</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Red x (✗) instead of green checkmark (✔)</td>
<td>All three ROs identified this failure</td>
</tr>
<tr>
<td>2</td>
<td>Green checkmark (✔) instead of red x (✗)</td>
<td>None of the ROs identified this failure</td>
</tr>
<tr>
<td>3</td>
<td>CPS skips a step in ES-1.3</td>
<td>None of the ROs identified this failure</td>
</tr>
<tr>
<td>4</td>
<td>CPS skips a step in E-0</td>
<td>None of the ROs identified this failure</td>
</tr>
<tr>
<td>5</td>
<td>Total loss of CPS</td>
<td>All three ROs identified this failure and transitioned to paper procedures in &lt;2 minutes</td>
</tr>
</tbody>
</table>

Heart rate, heart rate variability, breathing rate, and skin temperature were calculated for four different time periods: initial Baseline, computer procedure system (CPS) use: CPS–before switch (before switching to paper procedure; henceforth, CPS–before switch), CPS–after switch (interim time between when CPS stopped working and operators attempted to troubleshoot), and Paper procedure. Power spectrum analysis was used to calculate the power spectral density of heart rate variability in low frequency band (0.04 – 0.15 Hz) and high frequency band (0.15 – 0.4 Hz). Prior research indicates that increased heart rate, breathing rate, temperature and heart rate variability will typically be seen as inducing more workload.

Table 5 shows that both RO1 and RO3 had an upward trend in heart rate (HR) in the three conditions: CPS – before switch < CPS – after switch < paper procedure. However, RO2 had lowest HR in paper procedure, and also showed highest HR in baseline, which is unusual. This result might indicate a workload increase after the switch to the paper procedure. However, the physical actions involved in manipulating the paper procedure (e.g., page flipping) might have contributed to the HR increase.
Table 5: Heart rate (beat per minute)

<table>
<thead>
<tr>
<th>Crew Member</th>
<th>Baseline</th>
<th>CPS–Before switch</th>
<th>CPS–After switch</th>
<th>Paper procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>RO1</td>
<td>89.0</td>
<td>89.3</td>
<td>89.6</td>
<td>92.9</td>
</tr>
<tr>
<td>RO2</td>
<td>72.7</td>
<td>71.3</td>
<td>71.4</td>
<td>68.9</td>
</tr>
<tr>
<td>RO3</td>
<td>94.1</td>
<td>99.4</td>
<td>99.8</td>
<td>100.2</td>
</tr>
</tbody>
</table>

Heart rate variability (HRV), which is the beat-to-beat alterations of high rate, were captured for both the low frequency (LF) component, (0.04 to 0.15 Hz, which appears to be mediated by the vagus and cardiac sympathetic nerves) and high frequency (HF) component (0.18 to 0.4 Hz, which is synchronous with respiration) is presented in Table 6. Generally, lower HRV is indicative of greater sympathetic activity and, as a corollary, higher workload. In more sedentary tasks like CRO performance, much of this workload is mental effort and psychological stress.

Table 6: Heart rate variability (ms² power in low (LF) and high (HF) frequency bands)

<table>
<thead>
<tr>
<th>Crew Member</th>
<th>Baseline</th>
<th>CPS–Before switch</th>
<th>CPS–After switch</th>
<th>Paper procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>LF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RO1</td>
<td>717</td>
<td>929</td>
<td>3024</td>
<td>934</td>
</tr>
<tr>
<td>RO2</td>
<td>1733</td>
<td>977</td>
<td>953</td>
<td>926</td>
</tr>
<tr>
<td>RO3</td>
<td>1237</td>
<td>577</td>
<td>553</td>
<td>516</td>
</tr>
<tr>
<td>HF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RO1</td>
<td>86</td>
<td>194</td>
<td>297</td>
<td>304</td>
</tr>
<tr>
<td>RO2</td>
<td>433</td>
<td>93</td>
<td>106</td>
<td>82</td>
</tr>
<tr>
<td>RO3</td>
<td>2598</td>
<td>341</td>
<td>396</td>
<td>262</td>
</tr>
</tbody>
</table>

Artifacts in R-R interval recording could have influenced the results, notably for RO1 after the switch and possibly for RO3 during baseline. The R-R interval is the time elapsed between two consecutive R waves in the electrocardiogram or from the peak of one QRS complex to the next. An algorithm to detect and remove additional artifact is under development. For RO2 and RO3, HRV on HF indicated that workload might be highest in the CPS–after switch condition. This is corroborated by the modest increase in absolute heart rate after the switch to paper procedures (Table 5). Both workload and speech can influence average breathing rate (Table 7). In the current study, where communication was not controlled, breathing rate did not appear to be a reliable indicator of either variable. All the participants showed a steady increase in skin temperature across the experiment process (Table 8).

Table 7: Breathing rate (breaths per minute)

<table>
<thead>
<tr>
<th>Crew Member</th>
<th>Baseline</th>
<th>CPS–Before switch</th>
<th>CPS–After switch</th>
<th>Paper procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>RO1</td>
<td>12.2</td>
<td>13.0</td>
<td>15.5</td>
<td>16.7</td>
</tr>
<tr>
<td>RO2</td>
<td>16.9</td>
<td>15.5</td>
<td>13.5</td>
<td>15.3</td>
</tr>
<tr>
<td>RO3</td>
<td>19.9</td>
<td>14.8</td>
<td>12.9</td>
<td>14.7</td>
</tr>
</tbody>
</table>
Table 8: Skin temperature (degree Celsius)

<table>
<thead>
<tr>
<th>Crew Member</th>
<th>Baseline</th>
<th>CPS – before switch</th>
<th>CPS – after switch</th>
<th>Paper procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>RO1</td>
<td>31.8</td>
<td>33.0</td>
<td>34.2</td>
<td>34.4</td>
</tr>
<tr>
<td>RO2</td>
<td>33.2</td>
<td>34.9</td>
<td>35.7</td>
<td>35.8</td>
</tr>
<tr>
<td>RO3</td>
<td>31.2</td>
<td>32.9</td>
<td>34.3</td>
<td>34.4</td>
</tr>
</tbody>
</table>

This 3-participant one simulation study was too small to identify whether the physiological measures would be useful in future research. However, there was some suggestion that HRV may be the most useful as was found in our literature review. Future research using physiological measures should include complimentary subjective measures and observational measures for a comprehensive view of the construct of interest (e.g., workload).

Sequoyah Study

At one visit to Sequoyah NPP, we observed and collected psychological (not physiological) data during a simulated scenario where the plant experienced an emergent event. A total of four operators and two simulation instructors participated in this effort. The simulation instructor explained the occurrence of the events as they happened and interpreted what the operators were doing in response to events. We collected performance data included NASA-TLX, performance measure by scenario instructors, and physiologic data from four operators. The scenario lasted approximately two hours. The simulation director was unwilling to let us conduct the study during regular training scenarios.

Challenges

Our milestone to conduct multiple simulation-based studies at INL and/or at NPPs was not achieved principally due to ongoing changes in INL’s Benefits and First of a Kind (i.e., Control Room Modernization) initiatives and INL’s evolving relationship with the Palo Verde NGS. Our original plan was to participate in the proposed Benefits controlled simulation evaluation at INL of next-generation overview displays, starting in the summer of 2016. We were actively involved in the design and planning for the first weeklong simulation study of a Palo Verde crew (with the original plan to follow this with weeklong studies of at least two additional Palo Verde crews). However, due to technical difficulties at INL in programming their simulator to emulate the Palo Verde control room, in completing and integrating the overview display software, and in developing the scenarios required, these studies were postponed, first to the late Fall 2016 and then indefinitely.

As a fall back, we were relying on being able to participate in the First of a Kind workshops with Palo Verde, but due to several factors outside our control, we were not allowed to participate in or study these simulations. Chief among those factors was that INL’s negotiated legal agreement with Palo Verde included provisions to protect Palo Verde proprietary information. It was determined by both Palo Verde and INL project leadership (and lawyers) that Vanderbilt personnel were not covered by the agreement and could not participate in any of these ‘workshops.’ This left us with no way to collect the data we needed in time to complete this milestone before the end of Year 2. Further, during discussions with INL in Year 3 made it increasingly clear that we would be unable to achieve our simulation objectives at INL.
In Year 3, we pursued a three-pronged strategy for achieving our Aims of simulation evaluation: 1) On-site simulation study at Palo Verde NGS in collaboration with the First of a Kind Control Room Modernization project which was planning two on-site workshops) and/or the Benefits project; 2) Collaborative simulation study with Halden at NRC Chattanooga and/or South Texas NGS; and/or 3) On-site simulation studies at other NPPs (e.g., Sequoyah, North Anna, Watts Bar) during already scheduled training simulation sessions. Disadvantages of the third approach included inability to control the nature of the scenarios and to measure all of variables relevant to individual operator and system performance.

By the end of Year 3, we were only able to make limited progress using the third alternative approach. We conducted one ‘on site’ study at the Sequoyah NPP’s simulator, using an already planned training scenario, albeit with much hesitation from the utility. We were also able to collect data in a NPP simulator at the NRC Chattanooga facility, albeit not with active licensed CROs. Other attempts to negotiate studies or data collection with Watts Bar, North Anna, and Sequoyah were unsuccessful.

Our analysis was limited to the data that we were able to collect and will likely not be publishable in a journal article format and more studies would need to be done. We are exploring alternatives to present this in a meaningful way.

**Objective 4. Data fusion model development and verification**

In this task, a systematic information fusion framework was investigated to integrate heterogenous data from multiple sources and assess the operator performance in eliminating the malfunction events quantitatively. This objective required access to a rich multivariate dataset to develop a suitable data fusion model to integrate CRO performance and system performance in hybrid control rooms. Because of delays in the conduct of our proposal simulation trial (which could have generated these data), we had to obtain useful data from another source. Dr. Nathan Lau, our Virginia Tech collaborator, facilitated access to data from a previous nuclear simulation study he performed at the State of Virginia’s Center for Advanced Engineering and Research (CAER, Forrest, VA).

To facilitate this effort, we executed a formal collaboration and data use agreement with Bob Bailey, Executive Director of CAER. Due to institutional delays in executing this agreement, the applicable data only became available at the very end of Year 2. Then, Dr. Mahadevan (Aim 4 leader) and one of his graduate students, Xiaoge Zhang visited CAER along with Dr. Lau and his graduate student to developed a study plan. Dr. Lau familiarized the Vanderbilt engineering team with the structure and content of the available data.

These data were collected from a full scope, human-in-the-loop simulation previously conducted by Dr. Lau. The experiment employed a full-scale Generic Pressurized Water Reactor (GPWR) simulator in the CAER control room research facility and recruited nine previously licensed operators to form three crews. Each crew included the positions of unit supervisor (US), reactor-side operator (RO) and balance-of-plant-side operator (BOP). Ten scenarios were developed to test the performance of each group. Each experimental scenario consisted of two to four malfunction events.

The available experimental data, derived from multiple sources (e.g., test data, expert opinion, operational data, legacy system data, and model-based simulations), included: scenario
characteristics, eye tracking data, physiological data (i.e., skin conductance response and respiration), expert-rated task performance (i.e., Operator Performance Assessment System (OPAS), self-rated task performance, subjective workload ratings (the Halden Task Complexity Scale), situation awareness (SA) (the Process Overview Measure), and subjective SA confidence ratings.

The available heterogeneous data was in different formats and of variable fidelity thereby requiring appreciable data extraction and cleaning. A Bayesian network machine learning approach was applied for the fusion of these myriad types of information. Both discrete and continuous quantities were incorporated, with discrete variables representing categorical data. The resulting Bayesian network model was trained to predict the system and CRO team performance variables of interest. The model was verified and validated.

**Figure 3.** The flowchart of the proposed information fusion approach

Each of the aforementioned data type is in a different format, and an information fusion approach was developed for evaluating the crew team performance in a quantitative manner. Figure 3 summarizes the four major steps of the developed method. First, we extracted context-level characteristics of each scenario from the relevant documents, (e.g., scenario reports). These data provide scenario-wide features, including scenario difficulty, task complexity, and workload. The linguistic ratings on participants' situation awareness were converted into numerical scores by comparing the responses of the operators to those provided by the simulator operator and incorporating the operator confidence ratings. The physiological data, e.g., skin conductivity (also known as electrodermal activity or EDA) and respiration data, were subjected to event-related analysis to identify each operator's physiological response to the injected malfunction event, thereby extracting physiological features corresponding to each event. The physiological features included skin conductance response level, response latency, skin conductance response rise time, and the mean of respiration effort, etc. Third, the research team
elucidated the eye-gaze-based temporal (i.e., average fixation duration), spatial (i.e., the pursuit distance, the saccadic amplitude), and composite features (i.e., average duration of fixation on the visited important regions, the ratio during which important regions were covered, and the number of clusters corresponding to the eye movement). Finally, correlation analysis was conducted to reduce the problem dimensions and form the final set of input variables. All the previously extracted numerical features served as inputs for the machine learning algorithm.

With respect to model output, expert-rated OPAS values with variable number of items on the performance of each team in eliminating malfunction events was transformed into a four-variable representation with each variable indicating the numerical counts of that rating value across all the expert-rated items in each malfunction event.

Eventually, the study yielded a total of 107 data points and 22 input variables after several dimension reduction operations (correlation analysis, rating value averaging) were performed. To improve the model prediction capability, the support vector machine model was integrated with bootstrap aggregating to build an ensemble of models for the four individual expert rating scores, from which the classification results from numerous models trained by randomly selected samples are combined. *While no single variable predicted operator performance, the fused model's predictions using independent verification data was very good (prediction accuracy of 75-83%).* Further, the ensemble model of the full heterogeneous dataset was superior to SVM models built with only physiological or eye-tracking data, and substantially better than simple correlation between performance outcome (i.e., OPAS) and any individual input features.

The results demonstrate that this data-driven information fusion approach has promise for estimating performance of NPP operators, and could be used to predict the range of crew performance when using different control room technologies, an important adjuvant to other assessment methods given many on-going modernization and technological advancements.

The methodology development and demonstration problem results are documented in detail in an archival journal article currently under peer review for publication, as below.

4. **PUBLICATIONS & PRODUCTS**

**Publications:**


**Websites and Software Tools**

As part of our literature review, we created a password-protected database for the literature that includes all of the study variables, data coding, and links to the PDF of each article. This website includes an interface portal for data collection and review which is linked to the backend database that contains all of the data.

**Networks or collaborations fostered**

We continue to foster collaboration with the Halden Reactor Project, assisting them in meeting their research objectives through contributions to qualitative and quantitative performance measures. We were also able to collect performance data from an NRC training study during a simulated scenario.

Our collaboration with Nathan Lau at Virginia Tech has extended to a new NEUP project to develop advance software visualization and data analytic tools for NPP outage management.
5. REFERENCES


