INL/EXT-14-33901
Revision 1

Light Water Reactor Sustainability Program

A Research Framework for Demonstrating Benefits of Advanced Control Room Technologies

December 2014

DOE Office of Nuclear Energy

INL/EXT-14-33901
Revision 1

Light Water Reactor Sustainability Program

A Research Framework for Demonstrating Benefits of Advanced Control Room Technologies

Katya Le Blanc, Ronald Boring, Jeffrey Joe, Bruce Hallbert, and Ken Thomas

December 2014

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov/lwrs

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

ABSTRACT

Control Room modernization is an important part of life extension for the existing light water reactor fleet. None of the 99 currently operating commercial nuclear power plants in the U.S. has completed a full-scale control room modernization to date. A full-scale modernization might, for example, entail replacement of all analog panels with digital workstations. Such modernizations have been undertaken successfully in upgrades in Europe and Asia, but the U.S. has yet to undertake a control room upgrade of this magnitude. Instead, nuclear power plant main control rooms for the existing commercial reactor fleet remain significantly analog, with only limited digital modernizations. Previous research under the U.S. Department of Energy's Light Water Reactor Sustainability Program has helped establish a systematic process for control room upgrades that support the transition to a hybrid control room. While the guidance developed to date helps streamline the process of modernization and reduce costs and uncertainty associated with introducing digital control technologies into an existing control room, these upgrades do not achieve the full potential of newer technologies that might otherwise enhance plant and operator performance. The aim of the control room benefits research presented here is to identify previously overlooked benefits of modernization, identify candidate technologies that may facilitate such benefits, and demonstrate these technologies through human factors research. This report serves as an outline for planned research on the benefits of greater modernization in the main control rooms of nuclear power plants.

ACKNOWLEDGEMENTS

The authors would like to express gratitude to Rachel Shirley and Thomas Ulrich for reviewing drafts of this report.

CONTENTS

FIGURES

TABLES

ACRONYMS

AMCR	Advanced Main Control Room
BARS	Behaviorally Anchored Rating Scales
BOP	Balance of Plant Operator
CBP	Computer-Based Procedure
CMCR	Conventional Main Control Room
CFR	Code of Federal Regulations
CRM	Control Room Modernization
CVCS	Chemical Volume and Control System
DCS	Distributed Control System
DOE	Department of Energy
EPRI	Electrical Power Research Institute
EdAD	Event-dependent Assistant Display
FY	Fiscal Year
HAMMLAB	Halden Man-Machine Laboratory
HSI	Human-System Interface
HSSL	Human Systems Simulation Laboratory
HUPESS	Human Performance Evaluation Support System
I&C	Instrumentation and Control
INL	Idaho National Laboratory
INPO	Institute for Nuclear Power Operators
IRD	Information Rich Design
KAIST	Korea Advanced Institute of Science and Technology
KPI	Key Performance Indicators
LWRS	Light Water Reactor Sustainability

LWR	Light Water Reactor
NASA	National Aeronautics and Space Administration
NEI	Nuclear Energy Institute
NPP	Nuclear Power Plant
NRC	Nuclear Regulatory Commission
O&M	Organization and management
OPAS	Operator Performance Assessment System
PBP	Paper-Based Procedure
PPD	Procedure Performance Display
PSOD	Procedure Selection and Overview Display
R&D	Research and Development
RO	Reactor Operator
SA	Situation Awareness
SACRI	Situation Awareness Control Room Inventory
SAGAT	Situation Awareness Global Assessment Technique
SART	Situation Awareness Rating Technique
SRO	Senior Reactor Operator
STA	Senior Technical Advisor
TBP	Task Based Displays
TLX	Task Load Index
U.S.	United States

A Research Framework for Demonstrating Benefits of Advanced Control Room Technologies

1. Introduction

This Research is a part of the United States (U.S.) Department of Energy (DOE) sponsored Light Water Reactor Sustainability (LWRS) Program conducted at Idaho National Laboratory (INL). The LWRS program is performed in close collaboration with industry research and development (R&D) programs, and provides the technical foundations for licensing and managing the long-term, safe, and economical operation of current nuclear power plants (NPPs). One of the primary missions of the LWRS program is to help the U.S. nuclear industry adopt new technologies and engineering solutions that facilitate the continued safe operation of the plants and extension of the current operating licenses.

Control Room modernization is an important part of life extension for the existing light water reactor (LWR) fleet. None of the 99 currently operating commercial NPPs in the U.S. has completed a full-scale control room modernization to date. A full-scale modernization might, for example, entail replacement of all analog panels with digital workstations. Such modernizations have been undertaken successfully in upgrades in Europe and Asia, but the U.S. has yet to undertake a control room upgrade of this magnitude. Such technology remains the sole province of new reactors such as the four AP1000 plants currently under construction in the U.S. Instead, NPP main control rooms for the existing commercial light water reactor fleet remain significantly analog, with little evidence of digital modernizations. There have, of course, been select upgrades in the U.S. such as behind-the-boards modernization of crucial sensors, wiring, and controls. Additionally, there are a number of like-for-like replacements of obsolete or worn out components on the control boards such as like-for-like annunciator system replacements. There have also been several distributed control system (DCS) replacements for systems such as turbine control, feedwater, or chemical and volume control. These upgraded components and systems have typically addressed an immediate need to replace equipment that is past its usable life. Such upgrades rarely represent an encompassing or systematic vision for control room modernization and instead address primarily matters of equipment obsolescence.

As noted in *EPRI TR-1010042* (Electrical Power Research Institute, 2005), control room upgrades are scarcely an all-or-nothing undertaking. While it may be viable for one plant in a regulated market to complete a full-scale digital upgrade, the cost, expertise, and time required for such an upgrade is significant. The downtime required to replace a sizeable portion of an existing main control room well exceeds the outage cycle of a plant. In a commercial electricity market such as the U.S., it is challenging to justify the lost revenue of taking the plant offline to modernize the control room. Control room modernization, such as the fully digital control rooms found in some chemical and process control facilities, does not significantly decrease the cost of operating the plant, nor does it necessarily increase the safety or reliability of the plant. A commercial NPP's operating license requires a prescribed crew complement, regardless of the underlying technology in the control room. Further, the plant already operates at extremely high safety and reliability margins, and gains through digitization are likely to be minimal. Thus, the modernization of the control room becomes a sunken cost to the private utility, and there is little perceived benefit to the effort and cost required to replace the control room.

A survey of 11 U.S. utilities conducted by the Idaho National Laboratory (INL; Joe et al., 2012) revealed that there is a general desire among utilities to replace the existing control room with a fully digital modernized control room.[a] . However, the reality is that most utilities will only achieve modernization in a stepwise fashion—gradually digitizing one system at a time and creating a hybrid analog-digital control room. Further the end-state modernization is likely to be a hybrid control room that leverages as much advanced technology as possible, but not a fully modernized control room. Of the utilities surveyed, 50% identified cost as a significant upgrade barrier to a fully digital control room, while 20% identified the regulatory process as a barrier. The U.S. commercial nuclear fleet presents a unique situation for upgrades: while there is desire to upgrade, practical constraints such as cost (primarily through lost revenue) and regulations prove formidable hurdles to the upgrade process.

Previous research under LWRS has helped establish a systematic process for control room upgrades that supports the transition to a hybrid control room (e.g., Boring et al., 2014; Boring and Joe, 2014; Hugo et al., 2013; Ulrich et al., 2014). There are limits to the processes outlined in these guidelines. While the guidance developed to date helps streamline the process of modernization and reduce costs and uncertainty associated with introducing digital control technologies into an existing control room, these upgrades do not achieve the full potential of newer technologies that might otherwise enhance plant and operator performance. The aim of the control room benefits research presented here is to identify previously overlooked benefits of modernization, identify candidate technologies that may facilitate such benefits, and demonstrate the benefits of these technologies through human factors research. This report serves as an outline for planned research on the benefits of greater modernization in the main control rooms of NPPs.

[a] It is important to note that the participants who were surveyed were mainly operations staff rather than management and leadership, so the opinions expressed in this survey may not reflect the plant's official position

2. Control Room Technologies

The control room technologies that can be employed to enhance plant performance and efficiency range from the immediate need to modernize existing analog systems by replacing them with digital instrumentation and controls (I&C) to the integration of advanced displays, controls, and operator support tools. This section discusses these technologies based on the general time-line in which they might be deployed.

2.1 Current State

The main control room of an NPP in the U.S. typically represents a stand-at-the-boards configuration. The boards have a bottom horizontal bench board, on which are primarily found the controls; a vertical panel, on which is found a variety of indicators and some controls; and a top panel, which is largely reserved for annunciators. Across the control room—often in a semicircular or L-shaped configuration—there are groupings of various components and systems divided into primary side functions such as cooling and reactor control and safety, and secondary side such as turbine control and electric supply. A crew of operators—at a minimum a reactor operator and a balance of plant operator—monitors and controls the plant by moving through different positions at the boards. Most of the I&C is analog, with the exception of some digital systems of varying vintages scattered across the boards. The crew is completed by the senior reactor operator, who reads procedures and directs activities from a central area in the control room. The shape of the control room complements the position of the senior reactor operator, encircling him or her and allowing him or her to monitor the entire front panels from one position. The senior reactor operator will often have operator workstations from which he or she can call up key plant parameters to the extent they are available from the plant process computer and other digital systems. A senior technical advisor aids the operators and senior reactor operator with tasks in the control room, while supervisors may monitor and advise from a strictly hands-off position from the back of the control room.

This control room configuration, initially developed in the 1950s, is largely invariant across the commercial fleet of plants in the U.S. Accompanying the vintage of the design is much of the analog I&C, which simply represents newer variants of older designs. The control boards offer a wealth of information, but the boards do not necessarily prioritize the display of information. In fact, the analog indicators are always displayed, such that the presentation of plant information occurs in a parallel fashion that is always available to the operators. As such, the operators must scan the boards for relevant information as they monitor the plant. Alarms typically do not present in a prioritized fashion, and the interconnected nature of plant systems results in multiple simultaneous alarms when there is a fault. This can result in alarm flooding from an operator perspective during a plant upset (Persensky et al., 2010), but it also means that the annunciators form patterns that operators learn to recognize quickly at a glance.

2.2 Immediate Technologies

The analog instrumentation and controls in these control rooms are no longer manufactured on a large scale. While there are specialty vendors who will repair or manufacture new parts to vintage specifications, these represent as-needed custom replacement parts rather than a ready supply of new components. Plants have stockpiled pedigreed parts to replace end-of-life components, but this cache has dwindled over the life of the plant, and the supply was never intended to suffice for extensions beyond the original 40-year licenses that are now achieved regularly. Maintaining the reliability of these aging or outdated components and finding suitable replacement parts is a significant undertaking at an established plant. The piecemeal nature of repairing and replacing such components will often serve as a driver to modernization, as cost and downtime can be minimized with newer systems. At some point in the operational life cycle of the plant, it is more expensive to maintain an aging system than it is to replace it. In such cases, modernization is largely an artifact of the high cost of maintaining the existing system.

In other cases, there are clear advantages of replacing an existing system with a digital equivalent. For example, it may be desirable as part of a plant power uprate and turbine refurbishment or replacement to upgrade the turbine control system in parallel. The digital turbine control system may drive previously unutilized functionality such as automatic synchronization to grid that improves overall reliability during certain plant evolutions or frees operators from labor-intensive manual tasks.

Finally, industry best practices may drive digital replacements of existing analog systems. For example, improvements in steam generator level control may afford plants safer operation by introducing a new median select for controlling the instrument channel, thereby making the system fault tolerant to instrument failures. Operational experience at similar plants, regulatory recommendations, and vendor prescribed upgrades may drive such upgrades.

These upgrades, as noted earlier, create a patchwork of new digital amid legacy analog systems on the control boards. Procedures are updated to reflect the changes, and operators are trained on the new systems in the simulator prior to deployment at the plant. In most cases, these changes do not require a full license modification per 10 *Code of Federal Regulations* (CFR) 50.59 because the changes to the boards may not represent more than a minimal increase in the likelihood or consequences of an accident, because the changes to the boards do not affect safety critical systems (such as the reactor protection system or engineered safety features) and do not introduce significant new functionality (such as automation) to the plant. License requirements become a limiting factor on modernization, however, since the utility may be reluctant to undertake a potentially lengthy license modification process that might be triggered by significant new technology or functionality in the control room. The apparent benefits of the updated system do not clearly justify the effort required to achieve a license modification.

The LWRS Control Room Modernization Pilot Project has been working with utilities to minimize the inter-system variability between digital systems installed in a control room (Boring et al., 2014; Boring and Joe, 2014; Hugo et al., 2013; Ulrich et al., 2014). As new digital replacement technologies have been introduced into control rooms, they may represent significant variability in the human-system interface (HSI) due to differences in the digital systems deployed (e.g., generational differences in digital systems or general stylistic differences between different DCS vendors) or even due to differences in the developer of the systems (e.g., inconsistency of implementation of HSIs within the same DCS platform). By developing a consistent process to be used across multiple system upgrades, it is anticipated that utilities will be able to standardize the digital HSIs as they are introduced in a gradual, stepwise fashion (Boring and Joe, 2014).

The digital systems targeted for these stepwise upgrades maintain much of the existing functionality of the control boards but display information on mounted displays on the control boards (so-called digital islands) with different input devices such as mouse, keyboard, trackpad, and/or touchscreen. While functionality is kept largely consistent between the old and new systems, the HSI is, of course, considerably different. Rather than present information in a parallel fashion, as was the case with the analog boards, the DCS uses windows and pages of information and control, and the operators must learn to navigate between information. The experience working with utilities on the design and evaluation of such DCS replacement systems reveals that operators generally perform tasks more efficiently using the new HSIs compared with the analog boards (Ulrich et al., 2014). Such upgrades are a definite success to the utility in that they help replace obsolete I&C with up-to-date technologies that increase reliability and maintainability. These types of upgrades remain like for like and do not generally represent significant functional enhancements to the plant or control room operations.

2.3 Near-Term Technologies

Replacement of aging and obsolete equipment is, in many cases, necessary to extend the life of existing reactors. Beyond that, there are many technologies that could enhance control room operations and increase the efficiency and cost competiveness of existing plants. Many of these technologies could be deployed without significantly changing the existing concept of operations, and would augment existing systems rather than replacing them. Further, these systems could likely be deployed without a significant cost because they are essentially an addition to the control room rather than a replacement of existing systems.

2.3.1 Overview Displays

One category of technologies that could be deployed in the near term is overview displays. Overview displays are typically designed to help operators keep a big picture awareness of what is going on in the control room, which means that they can serve to augment rather than modify or replace existing sources of information. Overview displays typically serve as information presentation only and do not include any plant control capability, which means there is minimal risk in deploying them. Overview displays can be divided into two categories: large screen overview displays (which may occupy and entire wall in the control room) and smaller screen overview displays (referred to here as overview displays).

2.3.1.1 Large Screen Overview Displays

Large Screen overview displays provide a large, central display intended to provide crews with an effective tool for shared situation awareness. Large screen overview displays have been employed in control rooms in the oil and gas industry, grid control rooms, and advanced NPP control rooms. Expected benefits of large screen displays are improved situation awareness and crew coordination (Braseth et al., 2009; Stubler & O'Hara, 1996).

2.3.1.2 Information Rich Design in Overview Displays

One large screen overview display design employs a set of design principles termed Information Rich Design (IRD; Braseth et al., 2009). IRD is characterized by several principles that allow large amounts of information to be displayed in ways that facilitate situation awareness. IRD is currently implemented in large screen overview displays, but the principles may be implemented on smaller screen displays as well. One of the main features of IRD is the use of the "Dull Screen Principle" (Braseth et al., 2009). The Dull Screen Principle is characterized by conservative use of saturated color, which is reserved for important signals like alarms. The static elements of the display are presented in shades of grey to minimize interference with the important and dynamic elements of the display

Another important principle in IRD is the use of graphical display elements. Careful design of displays can reduce the amount of cognitive effort (i.e., memorizing and calculating) that is necessary when compared with simply displaying a digital value. One of the major examples of the use of graphical display elements is the use of normalized, integrated mini trends. Mini trend plots are integrated into the displays instead of simply displaying the digital values. On the mini trend plot, the scale is normalized so that the top is the high-high alarm set point and the bottom is the low-low alarm set point. The center (on the y axis) of the plot is the normal operating value. The region between the low and low-low alarm set points is shaded with a low contrast color (the same is true for the high and high-high alarm set points). These plots are grouped so that the set of plots is perceived as a single object and that deviation in a single parameter is easily detected (see Figure 1 for an example).

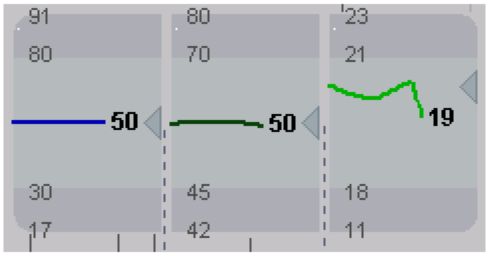

Figure 1. An Example of Mini Trend Plots in an IRD

Laarni et al. (2009) evaluated a large screen display prototype employing the IRD concepts at the Lovissa NPP in Finland. Aside from establishing that the operators liked the display and many of the display's features, one tangible result was that fault detection was significantly faster in one of the two tested scenarios. Large screen overview displays have been employed in control rooms in a variety of industries, but their effect on human performance has not been empirically tested exhaustively.

.

2.3.1.3 Overview Displays

While large screen overview displays may help to provide crews with shared situation awareness, many existing plants do not have enough room in their control rooms to accommodate the large screen overview displays. That does not mean that the concept of a central overview display should be abandoned. Many of the same concepts can be translated to smaller screens to create overview displays that can be used to get a general overview of the process or can be used to guide specific tasks. Smaller screen overview displays can provide a central location for information-dense presentation of important information. These displays can be designed to reduce operator cognitive workload by supporting pattern recognition in the same way that IRD principles do. They can also enhance efficiency by presenting relevant information in a centralized location so that operators do not have to "walk the boards" to find information.

2.3.2 Task Based Support Displays

Rather than providing a general overview, displays can be developed to support specific tasks. The task based approach provides operators with all information needed to perform pre-defined tasks as effectively and safely as possible. Procedure-based tasks are particularly suited to such an approach. At Halden, work in the HAMMLAB in support of task based displays (TBD) determined 3 types of displays were useful to support the task based display concept: the Procedure Selection and Overview Display (PSOD), the Procedure Performance Display (PPD), and the Event-dependent Assistance Display (EdA). Their research has demonstrated that many operators who used the TBDs preferred them over traditional displays (Braseth et al., 2009).

2.4 Farther Term Technologies

Many technologies that have potential to significantly enhance control room operations and efficiency would also require that existing systems be replaced or modified. These technologies may be more costly to implement than the technologies discussed above. They may also be subject to more regulatory scrutiny than the near term technologies. For these reasons, we have labeled them farther term technologies, but it should be understood that there are no technical hurdles that would prevent implementing these technologies today.

2.4.1 Advanced Alarm Systems

Alarm systems are an important aspect of any control room. In NPPs, alarms aid operators in monitoring tasks, decision making, action implementation, and verification of actions taken. Existing NPP control rooms typically utilize traditional alarm tiles. These tiles are generally sufficient to alert operators to important deviations in the process, but many problems with existing alarm systems have been identified (O'Hara et al., 2000). Among those difficulties are alarm floods, false alarms, and nuisance alarms. Alarm floods may increase operator workload and require the operator to sift through a large amount of irrelevant alarms to find relevant information. Alarm floods and false alarms may contribute to operators missing important information. Finally, nuisance and false alarms may activate during normal operation, making it difficult for an operator to determine if plant status is normal or abnormal (O'Hara et al., 2000). Several methods for advanced alarm processing have been proposed, including:

- **Filtering**. Suppress or eliminate alarms that are irrelevant based on current operational conditions.

- **Prioritization.** Present information regarding the level of urgency related to the alarms and/or order alarms based on urgency.

- **Context-Sensitive (or State-Based) Alarms**. Suppress alarms based on current plant mode or state.

Advanced alarms may result in reduced operator workload, quicker and more accurate fault detection and diagnoses, and enhanced plant performance; however, these benefits have not been demonstrated empirically. Braun, Grimes, and Shaver (2011) conducted a literature review on alarm systems research and identified only a single study that investigated the effects of alarm design on human performance. That study was conducted by O'Hara, Brown, Hallbert, Skraning, Persensky, and Wachtel (2000) and it compared the effects of alarm display types and availability of alarm information. The study found that a reduction in nuisance alarms alone was not sufficient to improve performance, and that a reduction in both nuisance and reduction of redundant alarms produced superior performance. Advanced alarm systems may reduce operator workload, increase efficiency and enhance diagnosis. Although there is a large amount of work on the technical aspects of advanced alarm design (See Braun, Grimes, & Shaver, 2011 for a review), there is less work demonstrating the effectiveness of those designs.

2.4.2 Computer-Based Procedures

Normal, abnormal, and emergency operations in an NPP control room are guided by procedures. Operators in the existing fleet of LWRs rely on paper-based procedures (PBPs) to provide step-by–step guidance to conduct almost all control room activities. Many researchers have documented the limitations of these paper-based procedure (PBP) systems, which include:

- Operators frequently have to manage multiple procedures at one time. This leads to navigational difficulties and high cognitive workload associated with keeping track of which procedure the crew is in and where the crew is within the procedure (Converse, 1995).

- Keeping place within the procedure has to be done manually, which may increase workload and lead to errors such as performing steps out of sequence (Fink et al., 2009; Le Blanc & Oxstrand, 2012).

- The information presented in the procedures is static and does not necessarily reflect actual plant conditions (Fink et al., 2009).

- Cautions and warnings may not be applicable to all systems states (Fink et al., 2009; O'Hara et al., 2000).

Many of these limitations can be easily addressed with computer based-procedures (CBPs). The concept of CBPs has been around since the 1980s. Early CBP systems include COMPRO (developed by Westinghouse Electric Corporation), COPMA (developed by the Halden Reactor Project), and the computerized procedures for the French N4 design. CBPs are in use or planned to be in use in many advanced plants internationally (Fink et al., 2009); however, at this time LWRs in the U.S. still use PBPs.

CBPs can be designed to employ a range of functionality, from static paper-like electronic documents to dynamic computer-based instructions with advanced capabilities. Several standards have characterized different types of CBPs based on the functionality provided by the CBP, including IEC 62646, "NPPs - Control rooms - Computer based procedures," IEEE (2009), and EPRI 1015313 (2010). The classification schemes typically start at the lowest level with a digital replica of the PBP and end at the highest level with a CBP equipped to execute automated sequences of procedural actions. Intermediate levels include operator support capabilities such as linking to supplemental information, links to soft controls that reside in control room displays, embedded process data displays, and automatic evaluation of procedure logic. The level of operator support afforded by the CBP has important implications for how the CBP system can help to address challenges of PBPs and for how the CBP systems may affect the operator's roles and responsibilities in the control room.

Several studies have sought to compare performance using CBPs to performance using PBPs, and have yielded mixed results. Converse (1995) compared error-rate, workload, and completion times between CBPs and PBPs in a simulated control room task and found that CBPs led to fewer errors but longer completions times. It is important to note that in this study, Converse defined looking ahead and back in the paper procedure (looking ahead was not possible with the CBP) as an error. Looking ahead in procedures is a desirable behavior; therefore, defining it as an error may have artificially inflated the PBP error rate and challenged the validity of these results. Chung, Daiwan, and Kim (2002) compared crew performance using CBPs and PBPs in a traditional and advanced control room and found that CBPs changed crew communications compared to CBPs, but did not find any other tangible performance differences. Roth and O'Hara (2002) assessed a crew's ability to use CBPs and transition to paper-based backups. They also investigated crew communication. They found that crews were able to use the CBPs and were able to transition effectively to the back-ups. They also identified a significant breakdown in crew communication using the CBP, indicating that CBPs may have a negative effect on crew communication. Lee, Hwang, and Wang (2005) compared CBPs that were either embedded into the HSI or presented as separate systems. They found that embedded procedures led to faster performance times than separate procedures, indicating that it might be important to integrate CBPs into the HSI to realize the full benefits of CBPs. Huang and Hwang (2009) compared CBPs to PBPs and measured error rates, performance time, workload, team performance, and situation awareness and discovered that CBPs enhance performance based on error rate, but situation awareness and workload did not differ between CBPs and PBPs. They also found that smaller crews (i.e., 1 person) were slower with CBPs than larger crews (i.e., 2-3 people).

There is some evidence that CBPs may enhance control room operation (Converse, 1995; Huang & Hwang, 2009), but many of the proposed benefits such as reduced workload and increased efficiency have not been demonstrated empirically. In fact, studies have indicated that CBPs may reduce efficiency by increasing completion times (Converse, 1995). Although the potential benefits of CBPs are compelling, research has failed to provide sufficient evidence that those benefits would be realized in an operational setting.

2.4.3 Predictive Displays

Predictive displays can use models or faster-than-real-time simulation to predict the trajectory of process parameters and present them to operators. Predictive displays were initially developed to aid telerobot operators (Noyes & Sheridan, 1984). Predictive displays have been found to aid operators in a variety of domains including aviation (Wickens, 2003), marine applications (van Breda, 1999), and process control (Yin et al., 2014). Anticipated benefits of predictive displays are a reduction in unnecessary operation and early detection and prevention of system failures (Yin et al., 2014). Many studies have demonstrated that predictive displays can enhance performance (Wickens, 2003; Yin et al., 2014; van Breda, 1999), but these benefits have not been demonstrated in an NPP control room.

2.4.4 Plant Automation

Upgrades to digital I&C may enable the use of advanced automation to carry out control room functions that are currently conducted manually. Many of the control room technologies discussed in this section inherently result in increased automation. For example, alarm processing may automate aspects of plant monitoring. CBPs may automate information gathering by providing embedded displays, and may even automated series of actions specified on the procedure. With well-designed, reliable automation, system performance is typically enhanced, because the automatic systems can perform many well defined routine tasks better than human operators can do manually. However, it is important to note that automaton may contribute to failure of the human-system if the automation fails due to a decrease in operator monitoring (Wickens, et al., 2010; Smith & Jameison, 2012; Manzay, Reichenbach, & Onnasch, 2008; Onnasch et al., 2013). The present research should establish the conditions under which plant automation provides a benefit.

3. Benefits of Control Room Technologies

The potential benefits of the advanced control room technologies discussed in Section 2 are numerous. Generally, the expected benefits include enhanced situation awareness, enhanced plant and operator performance, enhanced team performance, and enhanced efficiency. These benefits may be compelling to a human factors practitioner, but it is important to describe how these human factors benefits can further translate into tangible benefits for industry.

Formulating a value chain is one way to demonstrate how proposed technologies benefit the commercial nuclear industry. Rouse and Boff (2003) describe the value chain as a valuable construct to help understand how to tie initial costs to the expected return on investments. They write, "More specifically, it is quite helpful to consider the value chain from investments (or costs), to products, to benefits, to stakeholders, to utility of benefits, to willingness to pay, and finally to returns on investments" (pg. 641). In translating this generic value chain and applying it to this research problem, the value chain becomes the following questions:

1. What can operators do with this new technology that they could not do before? That is, by investing in these products, what are the expected technical benefits to operators and the plant?
2. What would that result be in terms of plant outcomes?
3. How would this show up in business or key performance indicators (KPIs)?

With respect to the first question, there is a range of technical benefits to the operators with the introduction of new control room technologies. They include reducing operator workload and mental burden. In NPP control rooms, the highest mental burden on the operators is typically during critical situations, and these technologies have the ability to reduce workload during these critical situations. The inclusion of technologies that automate routine activities also benefits operators. Automating routine sequences of activities eliminates the tedious, switch-by-switch control the operators currently perform, which then allows them to command the plant at a higher level rather than being weighed down in the tedium of getting systems aligned. Another technical benefit of these technologies is that they can help the operator integrate plant information when diagnosing plant conditions. This is particularly useful when the plant is experiencing abnormal and/or emergency conditions. These technologies can also provide trending information on key plant parameters that provide advanced notification to the operator, prior to any alarm set point, thereby assisting operators with their task of diagnosing plant conditions. Similarly these trends can be used to forecast the future state of the plant (via extrapolation of past and current plant conditions), providing additional time for the operator to address issues. It is also possible to extrapolate past and current sensor data, based on system models and energy balances, to create additional "virtual" sensors that further enhance the operator's understanding of plant conditions. In terms of improving the conduct of operations, the introduction of technologies that improve procedure implementation can provide numerous technical benefits, such as ensuring that operators correctly transition from different plant procedures based on plant conditions by validating the entry conditions, and automatically presenting the correct procedures to the operator. These technologies can also improve the communication interface between control room operators, plant support (i.e., field operators), and management systems. Finally, all of these technologies are designed with the philosophy of providing richer information in a more intuitive manner that improves the operator's comprehension of changing plant conditions. Table 1 summarizes the benefits described in Section 2.

Table 1. Summary of Benefits of Control Room Technologies

Technology	**Benefits**
Overview Displays	• Reduced workload (physical and cognitive) • Enhanced SA • Enhanced detection of off-normal conditions • Enhanced crew coordination
Task Based Support Displays	• Enhanced task performance • Reduced workload • Increased task efficiency
Advanced Alarm Systems	• Reduced Workload • Enhanced diagnosis • Increased efficiency
Computer Based Procedures	• Enhanced performance • Reduced Errors • Enhanced efficiency
Predictive Displays	• Reduced unnecessary operation • Early detection of system failures • Enhanced plant performance
Plant automation	• Enhanced System Performance • Reduced operator workload

With respect to the second question in our value chain (i.e., what would that result be in terms of plant outcomes?), the technical benefits described above would result in fewer safety challenges as operators are better able to detect and correctly diagnose off-normal conditions. Operators would be able to more quickly respond to plant transients, thereby minimizing the severity of the plant's deviations from normal operating parameters. The inclusion of these advanced monitoring and diagnosis technologies allows operators to focus on other ancillary duties with less concern that they are not maintaining adequate vigilance over the plant's state, and could lead to some reduction in the number of operations support staff (i.e., field operators). These control room technologies can also provide indications of the plant's state to field operators and other support staff outside of the control room, thereby facilitating the dissemination of information to key support personnel without burdening control room operators with the task of communicating this information. Overall, these technologies will result in fewer time-critical actions by both control and field operators.

With respect to how the results mentioned previously would manifest themselves in business or KPIs, the connection is fairly straightforward. There are a number of standard KPIs that would clearly be affected by the aforementioned technological enhancements, including a higher capacity factor, reduced forced loss rate, reduced Operations and Maintenance (O&M) costs, reduced radiation exposure to plant staff, and improved ratings from the U.S. Nuclear Regulatory Commission (NRC) and Institute for Nuclear Power Operations (INPO). Table 2 summarizes this value chain presented for advanced control room technologies.

Table 2. The Value Chain for Advanced Control Room Technologies

Technical Benefits
• Reduce operator workload and mental burden. • Automate sequences of activities to reduce tedious manual control and associated human error. • Assist the operator in integrating plant information to make diagnosis of plant upsets. • Provide operators with early warnings of trends by validating them far below the alarm setpoint. This buys considerable time to deal with conditions. • Provide accurate forecasts of where the plant will be at future times based on the extrapolation of plant conditions and the expected response of plant systems. • Provide virtual sensors based on system models and energy balances. This will greatly augment the data set available to the operators. • Ensure correct transition to plant procedures based on plant conditions. Automatically present the procedures to the operator, validating entry conditions. • Provide a seamless interface to plant support work and management systems. • Provide richer information in graphical forms that increases the operators' rate of acquiring an understanding of changing plant conditions.
Technical Benefits lead to Improved Plant Performance:
• Fewer safety challenges due to operator's failure to detect off-normal conditions. • Quicker responses to plant transients resulting in less severe plant deviations and better outcomes. • Allow operators to perform ancillary duties without concern on ineffective plant monitoring. Could allow reduction of some Operations support staff. • Greater throughput of support work when Operations can be more responsive and certain plant work activities can proceed without control room interaction. • Fewer time-critical operator actions.
Improved Plant Performance Improves KPIs
• Higher Capacity Factor • Reduced Forced Loss Rate • Reduced O&M Cost • Reduced Dose • Improved Regulatory Ratings

4. Research Plan

Very few of the proposed benefits of advanced control room technologies have been empirically demonstrated. An independent evaluation of the potential benefits of these technologies will provide a basis for industry to decide whether they should adopt these technologies and how to prioritize their adoption. This section describes the research approach that INL researchers will take to investigate these benefits.

4.1 Approach: Human in the Loop

We propose using proved human factors techniques to evaluate the benefits of the new technologies available for control room upgrades. The key to human factors research, as described by Meister (2004), is the measurement of how human behavior is affected by technologies and uniquely, the way human behavior can transform that technology. The primary goal of human factors research is to refine technology—to optimize it to compensate for the limitations and accentuate the strengths of its human users. Specifically, in the context of nuclear technology, the goal of human factors is to design the technology to maximize operator performance. Operators are not just trained or optimized to use technologies put before them; rather, through human factors, technologies are optimized to conform to the operators and to maximize their performance while using that technology. Such optimization of technology requires involvement of the actual users during the design of the system. Although we are not designing systems in this study, many of the techniques used during system design can also be used to demonstrate the benefits of a developed technology, or to compare and contrast two systems.

Human factors features an extensive set of methods (e.g., Stanton et al., 2013), but the commonality across approaches and methods entails a process of verifying and validating operator-technology interactions (or, more broadly, simply the HSI). Verification, as defined in Boring et al. (2014), involves "checking the HSI against an existing human factors standard like NUREG-0700 (U.S. NRC, 2002), while validation requires checking the performance of the system and operators according to desired performance" (pp. 1909-1910). Where standards and guidance documents readily exist, human factors experts can *verify* the HSI against desirable and understood qualities of performance. For example, the IRD Dull Screen principle is a response to known limitations of operator performance in the face of multicolored indicators in a digital HSI (Ulrich et al., 2012). Where standards and guidance documents do not exist or lack the detail to predict operator-technology interactions, the HSI is *validated*. A validation involves empirically observing and documenting operator performance in the use of an HSI. The standard way of validating an HSI in human factors is through a human-in-the-loop study. Within nuclear power, this validation becomes an operator-in-the-loop study.

Ostensibly, an operator-in-the-loop study should simply be a matter of exposing operators to the new HSI they will use. In reality, there are significant considerations that must be addressed to perform such a study correctly. These considerations include:

- *System technology:* The technology being introduced into the control room is one of the most important factors to shape the operator-in-the-loop study. As discussed earlier in this report, control room modernization can take many forms, including the introduction of significant new digital technologies into today's largely analog control panels. New technologies may include new forms of visualization (e.g., windowed screens vs. traditional always visible indicators), new types of input (e.g., touchscreen vs. legacy physical controls), new functionality (e.g., new sensors and new control logic), and even new roles for the operators (e.g., monitoring of automated processing vs. manual operations). New technology may bring about significant changes to the conduct of operations, and it is desirable to document and validate these changes. The introduction of multiple changes to the conduct of operations (e.g., new automation possible through new sensors plus control using touchscreens) may make it difficult to disambiguate the effects of operator performance, because the concurrent changes serve to confound the results. In such cases, it is desirable to explore the individual effects of changes in addition to the overall effect in order to have a clear model of how each change affects operator performance. A series of studies may be prescribed to identify the individual effects of each change in addition to the composite effect.

- *System fidelity:* This topic addresses the degree to which the HSI being used is representative of the system that is or will be deployed. In some cases, it may be desirable to test surrogate technologies (e.g., testing operator response to a system deployed in a non-nuclear process control setting). Alternately, it may be desirable to test a mockup of a system to gain early operator feedback on the design concept. It may be desirable to test a fully functional prototype or to conduct the equivalent of a factory acceptance test on a completed system. The degree to which the HSI represents the final deployed system may be varied systematically as part of the study design.

- *Design stage:* The fidelity of the system will change as the system is developed. As part of an overall user centered design process, it is desirable to test different iterations of the HSI at different stages of completion (Boring et al., 2014). In this manner, the HSI design may be refined based on operator performance and feedback. Human factors is most effective when it can provide early, formative input into the design prior to deployment. Formative studies most effectively allow the operators to optimize the design of the HSI, whereas later stage (so-called summative studies) are most effective for confirming the effectiveness of a human-factored HSI. As the system is developed, it is assumed it will converge on the final deployed system and will become a full-scope HSI, meaning it encompasses the full functionality of the system. Thus, there is generally a relationship between the design stage and the fidelity of the system being evaluated—early stage design mockups will tend to be of lower fidelity than latter stage designs ready for deployment at the plant.

- *Environmental fidelity:* This topic is also related to system fidelity but encompasses the context of the HSI. For example, an advanced chemical and volume control system (e.g., see Thomas et al., 2013) may be developed as a standalone system. This system can be interfaced with a plant or a plant simulator model, but it also exists independent of the context of the plant. It is possible to design, develop, and test that HSI outside the context of its operation in the plant and outside its interaction with other systems in the plant. This approach can be very effective for developing the proof of concept for the system. However, it is also necessary to evaluate the HSI in the context of its eventual operation in the plant. For example, testing the HSI in context of the plant panels is a requirement of integrated system validation in NUREG-0711,(U.S. NRC, 2013). Validating the environmental fidelity of the HSI is also necessary for external validity—the extent to which the findings from a study can be generalized to the deployed system. The importance of conducting human factors studies of the system in the context of its use is one of the important rationales for the creation of the Human Systems Simulation Laboratory (HSSL; Boring et al., 2012 and 2013). The HSSL serves as a full-scale testbed for operator-in-the-loop studies in which new control room technologies can be validated amid existing plant systems and in the context of the overall main control room.

- *Location of study:* Every commercial NPP has a full-scope, full-scale control room simulator at its disposal. While this may seem the logical place to host an operator-in-the-loop study, there are limitations in terms of the availability of such facilities (whose primary purpose is the mandatory training of crews). Additionally, there are restrictions on the types of changes that can be carried out in these facilities, as they must stand at the ready for training at the conclusion of the study. A simulator at a plant that features the exact hardware mimics of the instrumentation and controls of the actual main control room may not lend itself to physical reconfiguration, which involves removing devices and, in the case of inserting digital displays, cutting holes in the control boards to mount the displays. To allow a more flexible environment for studies, research simulators have been developed, which can be fully customized to incorporate the technology required for a human factors study (Boring, 2011). Prominent examples of research simulators for operator-in-the-loop studies include the Halden Man-Machine Laboratory (HAMMLAB) at the Halden Reactor Project in Norway and the HSSL at INL.

- *Nature of the users*: It is assumed that the operators used in a study on control room modernization are licensed reactor operators. There are several possible variants on this assumption. For example, if it is desirable to perform evaluations using a sufficient crew sample size to yield statistically significant results, it may not be practical to draw enough operating crews from a single plant, especially a single unit plant. Thus, it may be desirable to evaluate crews from similar plants, including crews not qualified on the specific plant model being used. Such would be the case when evaluating crews using the Generic Pressurized Water Reactor simulator model found in the HSSL. Limitations of this approach must be noted, including the possibility that the crew's lack of familiarity with some plant-specific configurations could influence the results of the study. In other cases, it may be desirable to enlist operators in training to gain perspective on future users of the system or recently retired operators to capitalize on the expertise of the most skilled crews. Using a variety of operators can help answer very different types of research questions.

  Another user consideration entails the extent to which it is feasible to use the same crews across multiple scenarios. Using the same crews—a so-called *within-subjects* experimental design—affords the advantage of maximizing the number of data points that can be collected from each crew, taking maximum advantage of each crew and the considerable expense and time required to use them in the study. On the other hand, reusing the crews introduces study confounds such as practice effects that can impact operator performance on the system being used. It may therefore be desirable not to reuse the crews across conditions and scenarios. Not reusing crews in similar scenarios is called a *between-subjects* experimental design. Its primary disadvantage, from the standpoint of operator-in-the-loop studies, is the need to recruit additional crews to participate in the study.

- *Types of measures:* Human factors is aligned to the field of psychology, which is crucial in establishing how to measure operator performance. Operator performance may be assessed observationally (e.g., recording and rating performance by a subject matter expert), objectively (e.g., data logging operator actions on the simulator), or subjectively (e.g., surveying operator impressions of the HSI or their task performance). Additional measures may be gathered to assess the operators during particular points during the study (e.g., physiological states such as heart rate during a plant upset), which can help establish the underlying causes of performance outcomes. Additional measures can pinpoint aspects of operator cognition while monitoring, making decisions, or carrying out actions (e.g., eye tracking may be used to determine where an operator is looking throughout the study, or subjective measures of workload may be administered to the operators). It should be noted that the purpose of such performance measures is not to identify deficiencies in the operators but rather deficiencies in the HSI as manifest in operator performance.

- *Goals of the study:* The goal of the study will logically shape the design of the study. If the goal is to shape the design of a new distributed control system in the control room, the study may be set up as a usability study that identifies how operators use the system and what aspects of the system could be improved. For such a study, simple objective and subjective measures will be sufficient for the evaluation. In contrast, if the goal of the study is to understand the mental model of how operators use, for example, procedures and create a new computer based procedure system based on that mental model, the measures collected would focus on capturing cognitive aspects of operator performance (Oxstrand and Le Blanc, 2012).

  In the context of control room modernization, one definite goal of the study is to compare the performance of operators using the conventional or existing analog control boards vs. the new digital HSI. In such cases, it is desirable to use a benchmark method in which operator performance is compared between the two systems. As discussed above, the participants in the study may participate in both the baseline and comparison study (i.e., a within-subjects experimental design), or the benchmark may enroll different crews in the two conditions (i.e., a between-subjects experimental design). A baseline study should be conducted using the existing control boards (typically at the host plant), and a replicate study is conducted using the new HSI (see Boring and Joe, 2014). Note, however, that when the new HSI represents a significant departure from the existing control boards, it may not always be possible to conduct an initial baseline study, because comparable baseline functionality may not be available.

- *Types of scenarios:* This consideration addresses what the operators will do during the study. The precedent for the studies is clearly the simulator training scenarios that operators routinely undergo for normal and off-normal plant conditions. The scenarios should ideally be developed under consultation with a qualified instructor at the plant in order to ensure the realistic trigger for and evolution of an event. Scenarios will often be developed that test operator performance under plant upset conditions, which are precipitated by a fault in a system or component. Depending on the goals of the study, it may be desirable to test the new HSI under a variety of operational conditions, including normal, abnormal, and emergency operations. Where the HSI extends the functionality of the plant for emergency operations and where the simulation supports such conditions, it may be desirable to evaluate operator performance for beyond design basis accidents.

The following sections will build on these considerations and explore the most important research measures in the context of the control room benefits project. These factors comprise the essential character of the operator-in-the-loop study. These factors also give considerable flexibility to the control room benefits project to evolve the human factors approach to address multiple facets of control room modernization as part of a systematic research program

4.2 Performance Measures

Evaluating the benefits of advanced control room technologies requires that the proposed technologies be demonstrated empirically in a realistic context. Section 4.1 describes the general approach to comparing existing technologies to the advanced control room technologies in simulator studies. This section describes the approach to measuring performance so that the benefits can be assessed empirically. Appendix A provides a summary of how the technologies map to the potential benefits, performance metrics, and the approach to scenario design.

4.2.1 General Approaches to Performance Measurement in Simulator Studies

There have been a few attempts to develop a comprehensive suite of performance measures for simulator studies. This section describes one of the most prominent standardized suites, the Human Performance Evaluation Support System (HUPESS). While HUPESS represents the latest development in comprehensive measurement frameworks for simulator studies, there is other noteworthy work that has been published on this subject. One example is EPRI's technical report 1010042 (2005). Section 3 provides guidance on methods and tools for collecting information from end users (e.g., operators). Additional guidance on the use of performance measures in full-scope simulator studies can also be obtained by reviewing Hallbert et al., (2014) and Hallbert, Sebok, and Morisseau (2000).

4.2.1.1 HUPESS

The Korea Advanced Institute of Science and Technology (KAIST) developed the HUPESS, which is a multi-tiered, computer assisted, suite of measures to assess operator and system performance in simulator studies (Ha & Seong, 2009). Figure 2 presents an overview of the HUPESS framework, and shows that it consists of both real-time objective assessments and post-test subjective assessments of operator performance by both expert observers and the operators themselves (i.e., self-evaluation). Both of these assessment activities are supported by a robust collection of technologies and computer systems, including a simulator server, an instructor station, eye-tracking systems, mobile and stationary systems to support expert and operator evaluations, and audio/video recording systems.

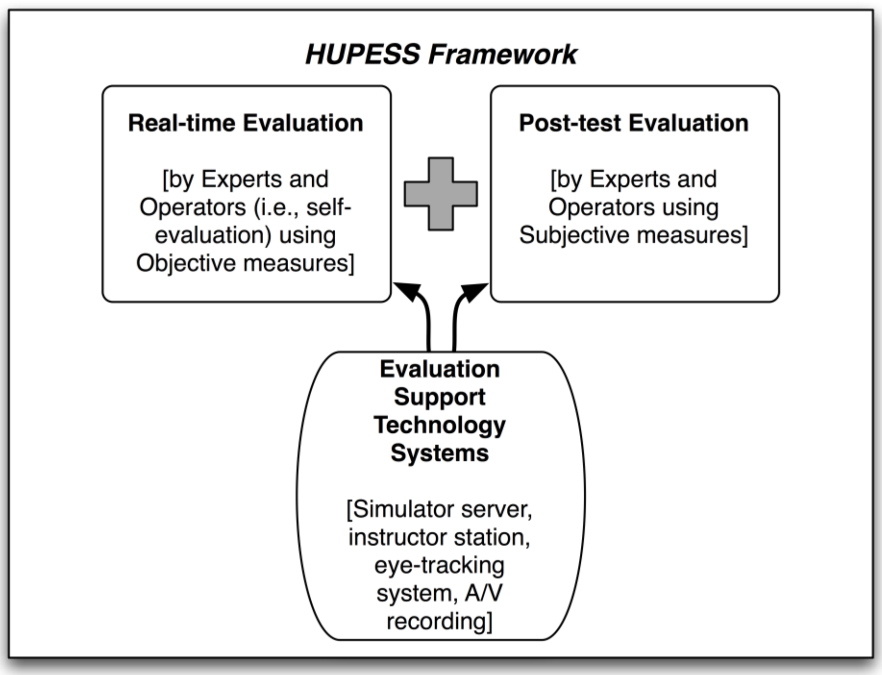

Figure 2. The HUPESS Framework

One important aspect of HUPESS to note is that, consistent with the previous discussion of establishing a value chain in Section 3, optimal human performance is not the "bottom line" measure in HUPESS. Rather, HUPESS defines optimal overall plant performance as the primary measure. Measures of operator performance, both objective and those that subjectively assess their cognitive state, are considered secondary performance measures in HUPESS that need to align with this primary measure.

From a process perspective, HUPESS has a three stage, step-by-step approach that ensures the systematic and experimentally valid collection of performance evaluation data from simulator experiments. The three stages are: Preparation, Evaluation, and Analysis. Preparation entails constructing scenarios that allow for the collection of operator and system performance data to test the experimental hypotheses. The scenarios that are constructed also need to be realistic and within the simulator's capabilities to simulate. Key human performance measures and plant performance/process parameters are also identified at this time, and their relative importance is determined and weighted accordingly. The Evaluation stage is when licensed operators are brought into the simulator and run through the scenarios developed during the preparation stage. As Figure 2 shows, as the operators work through the designed scenarios, both experts and operators perform real-time assessments of operator and system performance using objective measures. Operator performance, or personnel task performance, is primarily measured with the objective measures of completion time (using time tagging) and confirming the degree to which their performance conforms or deviates from a predetermined optimal solution path. Eye-tracking is also used as a secondary real-time measure of operator performance, and provides some insights into the operator's level of situation awareness and workload level. Aspects of teamwork are also assessed in real-time via observations using a time-tagged coding scheme for key teamwork behaviors (e.g., effective communication, awareness of when individuals are becoming overloaded and providing assistance, etc.). The Analysis stage occurs after the simulation scenarios have ended, and entails both the administration of subjective measures of performance and in-depth analysis of the data collected during the scenario. Subjective measures issued at this time include computerized questionnaires assessing the operator's situation awareness and workload. Computerized behaviorally anchored rating scales (BARS) are used to assess teamwork. The audio/video and eye-tracking recordings are also analyzed in more detail during this stage.

Overall, HUPESS is a coherent and well-organized approach to measuring human performance (in the context of overall system performance) in experimental simulator studies. It strikes a balance between a number of trade-offs, including the use of both objective and subjective measures, and the expediency of collecting real-time data versus the thoroughness of collecting and analyzing post-experimental data.

4.2.2 Specific Measures and Metrics for Performance

No standardized suite can account for every research question or experimental context; with the HUPESS framework as a guide, the researchers will select appropriate performance measures to demonstrate the specific benefits of the technologies during the types of studies described in Section 4.1. This section describes the performance measures (i.e., the general concepts of human and system performance) and metrics (the specific measurement tools) that may be used in this series of studies. Many of the metrics described are already incorporated into the HUPESS data collection framework.

4.2.2.1 Plant Performance

Many of the benefits outlined in Sections 2 and 3 can be summarized as various indicators of enhanced plant performance. Plant performance can be described as the combined performance of the human and the system. Therefore, in order to assess overall plant performance it is important to both assess system performance and evaluate operator performance. Plant performance can be measured by attempting to measure directly how well functions are executed by the human-system or by measuring operator performance and inferring the effect that operator performance would have on plant performance. It is important to note that while plant performance metrics and operator performance metrics are thought to be two aspects of the same thing, operator task performance and plant performance may be only moderately correlated (Jang, Park & Seong, 2012). It is therefore important to measure both to ensure that all relevant aspects of performance are captured. In addition, aspects of operator performance that do not necessarily have any observable behavioral consequences in a typical control room task (such as situation awareness and cognitive workload) are often measures in studies investigating control room performance. This section describes the measures and metrics that will be used to evaluate both system performance and human operator performance.

4.2.2.1.1 Discrepancy Scores

The main plant performance metric used in HUPESS is a discrepancy score. Ha and Seong (2009) describe a method to measure the discrepancy between the prescribed values of important plant parameters. The important parameters and the acceptable ranges of those parameters are defined by subject matter experts and are dependent on the specific scenario being tested. Once the parameters and ranges have been defined, the following formula is used to calculate a discrepancy score.

Discrepancy at time t, $D_i(t) = \begin{cases} \frac{X_i(t)-S_{Ui}}{M_i}, if\ X_i(t) > S_{Ui} \\ 0,\ if\ S_{Li} \leq X_i(t) \leq S_{Ui} \\ \frac{X_i(t)-S_{Li}}{M_i}, if\ X_i(t) < S_{Li} \end{cases}$

- S_{Li} = lower bound of parameter i
- S_{Ui} = upper bound of parameter i
- $X_i(t)$= value of parameter i at time t
- M_i = Mean value of parameter i during steady state

Average discrepancy for parameter i = $\sum_{t=0}^{T} \frac{D_i(t)}{T}$

Discrepancy scores can be used to calculate time outside of desired range or the maximum discrepancy in a given time period. The scores can also be computed for specific periods of time in the scenario to reflect different conditions. The calculation can also be modified to reflect the difference between an observed value and a range specified in a procedure or a maximum value (specified in technical specifications). The specific implementation of this measure will depend on the benefits and technologies being evaluated. For example, testing whether predictive displays enable the operation with a broader safety margin would require that the difference be calculated between the observed value and the maximum acceptable value.

4.2.2.1.2 Other Metrics for Plant Performance

There are many other ways to measure plant performance that are similar to discrepancy scores. An alternative way to assess plant performance based on monitoring important plant parameters was reported in O'Hara et al. (2000). In their metric, important plant parameters were identifies by process experts and then weighted based on importance. The values of the parameters were then compared to an "optimal" run of the simulation scenario, and relative weighted deviations were computed. Plant performance can be assessed in an analogous fashion by comparing parameters to technical specifications and determining the safety margin between observed performance and the technical specifications. These methods are very similar to discrepancy scores; however the discrepancy score metric is a more systematic and formalized approach. For this series of studies the primary plant performance measure will be discrepancy score. The score will be adapted or modified or adapted as needed to accommodate different research questions.

4.2.2.2 Operator Performance

Operator performance encompasses several perceptual, cognitive and physical activities. Operators must effectively monitor the plant, diagnose any issues, and plan and execute action plans. An approach to assessing operator performance must encompass all of these. This section describes objective and subjective methods for measuring operator task performance.

4.2.2.2.1.1 OPAS

Halden Reactor Project's Operator Performance Assessment System (OPAS) is a framework and methodology that analyzes and assesses operator performance in simulator experiments (Skraaning, 2004). OPAS was created out of the recognition that measuring human performance has many philosophical, methodological, and in the case of complex simulator studies, practical challenges as well. At the highest level, OPAS is designed to measure differences in system performance as a function of the quality of operator performance (e.g., quality of operator diagnosis, problem solving, etc.). OPAS is a computer assisted, hierarchically structured, real-time measurement system that assesses system and operator performance against predefined standards of performance and predetermined goals that are established when the simulator scenario is being formulated by subject matter experts in advance of the experiment. The subject matter experts decide or determine what the main goal is for a given scenario, and then further identify sub-goals that must be accomplished in order to achieve the main goal. Sub-goals are further divided into actions that the operator must perform to achieve the sub-goal. The operator actions and sub-goals are differentially weighted on a 5-point scale to reflect their importance in achieving the associated sub-goal (for operator actions) or main goal (for sub-goals).

The measures in OPAS focus on observable operator performance (i.e., behaviors). Assessment of cognitive processes and cognitive constructs, such as situation awareness and workload, are not a part of OPAS's measurement suite[b]. Rather, four primary measures, which align to the hierarchical structure described above, are used to assess observable operator behaviors. They are: performance ratio, relative importance, relative importance ratio, and the general performance score. The performance ratio is defined as the sum of the observed activities the operator performed (which are weighted) divided by the sum of all the operator activities (also weighted) that were predefined as the correct actions by the subject matter experts. The relative importance measure is the primary sub-goal measure, and is the weight of each individual sub-goal divided by the sum of weights for all sub-goals. The relative performance ratio is calculated by multiplying the results of the performance ratio calculation and the results of the relative importance calculation together. This relative performance ratio would be calculated for all sub-goals identified. The general performance score is then the sum of all of the sub-goal relative performance ratios previously calculated, multiplied by 100.

Skraaning (2004) notes that the results of the real-time performance measurement by OPAS can be further supplemented through post-experimental debriefs with the operators, and analyses of audio/video and eye-tracking recordings, though he also expressed caution with respect to relying heavily on retrospective self-reports from operators due the subjective nature of these kind of data.

Overall, OPAS is considered by many to be a significant methodological advancement in that its structured design and systematic implementation addresses a number of known shortcomings and

---

[b] Interestingly, problem solving is considered a behavior and is measured. Even though some may consider problem solving a cognitive activity, OPAS operationally defines it as an observable behavior, in that the experimenters have equipment in the observation gallery that allows them to see both where in 3-D space the operator is looking and the displays in that space that they operator is likely looking at. Thus, problem-solving behaviors are inferred from these data sources.

challenges with measuring human performance in complex simulator studies. The question of whether OPAS, in its efforts to be reliable and objective, also limited its usefulness by excluding measures of cognitive phenomena is a question that will continue to be debated among researchers in this field.

4.2.2.2.2 Expert Observation of Task Performance

The majority of research investigating control room performance uses expert observation of operator performance. OPAS represents a systematic and standardized approach to observationally measure operator performance, but there are other ways to measure operator performance using observation. In general these approaches define criteria for successful task completion and assess performance against those criteria. Several methods use objective observable behaviors as the basis for evaluating task performance such as predefined errors or deviations from the ideal sequence of actions (Roth & O'Hara, 2002; Xu et al., 2008; Huang & Hwang, 2009; Lois et al., 2009). Other methods use subjective qualitative methods such as an unguided expert evaluation that simply rates performance as satisfactory or unsatisfactory (Carvahlo, 2006; Jeffroy & Charron, 1997). The less formalized methods for observationally evaluating performance may be appropriate in contexts where a systematic approach like OPAS overspecifies the parameters for performance (e.g., in cases where there are multiple ways to solve the problem or the scenario develops in an unexpected way) or when research questions are more specific (e.g., if the benefit that the researchers want to demonstrate is simply fewer errors). Task performance will be assessed using OPAS and additional measures as needed.

4.2.2.2.3 Operator SA

Situation Awareness (SA) refers to an operator's awareness of what is going on (Endsley 1995). Although there are several models of SA (e.g., Klein, Moon, & Hoffman, 2006, Smith & Hancock, 1995; Bedny & Meister, 1999), the most prominent model of SA is Endsley's (1995). Endsley describes SA as an operator's ability to perceive relevant elements of the environment, comprehend what they mean, and predict their development over time. SA is a particularly important aspect of performance to measure in an NPP control room because a large portion of an operator's task is monitoring and detection. Further, advanced technologies can have unintended consequences for SA. For example, displays intended to make an operator's job easier and enhance SA by providing all relevant information in a central location (such as overview displays) may narrow an operator's attention, causing them to miss indications outside that display (this is often referred to as the keyhole effect). Increases in automation, such as procedure-based automation in a CBP system or automation built into an advanced control system might also affect SA (Endsley, 1995; Wickens & Hollands, 2000; Wright & Kaber, 2005). Finally, the proposed benefit of many of the advanced control room technologies is enhanced SA. In order to adequately assess the benefits (and potential human performance costs) of these technologies, SA should be measured in each of the studies.

4.2.2.2.3.1 SACRI

The Situation Awareness Control Room Inventory (SACRI) is an objective method for measuring SA in NPP control rooms (Hogg, Follesø, Strand-Volden, & Torralba, 1995). SACRI is adapted from a SA measurement technique developed for aviation called the situation awareness global assessment technique (SAGAT) which was developed by Endsley (2000). SACRI is administered through a freeze-probe questionnaire. That is, during a simulation scenario, the simulation is put in freeze mode and a questionnaire that targets the current state of important parameters and the development of those parameters is given to the operators. The important parameters that are targeted in the questionnaire are based on input from a subject matter expert, and the questions typically ask about the current value of the parameter, the development over the past and the future development.

4.2.2.2.3.2 SART

The Situational Awareness Rating Technique (SART; Taylor, 1990) is a questionnaire method for measuring SA that focuses on measuring operator knowledge in three areas (on a 7-point ordinal scale from low to high): demands on attention, supply of attentional resources, and understanding of the situation. SART is administered after the task is complete. SART and SACRI will be used in conjunction to measure SA in this research.

4.2.2.2.4 Operator Workload

Reduced workload (both physical and mental) is a proposed benefit of many of the control room technologies discussed in Section 2. Although workload is commonly used construct, there is no definitive approach to measuring workload.

4.2.2.2.4.1 NASA TLX

One of the simplest and most widely used metrics for operator workload is the National Aeronautic and Space Administration Task Load Index (NASA TLX). The NASA TLX (Hart & Staveland, 1988) is a subjective 6-item scale that is a widely used and validated scale for measuring workload after a task (Ha et el., 2007; Sebok, 2000; Huang et al.,2006; Converse, 1995; Hwang et al., 2008; Le Blanc Gertman, & Boring, 2010; Miller, 2001;. It was developed specifically for the aviation industry, though it has been used in hundreds of studies in a wide variety of fields, including many NPP control room studies (Le Blanc Gertman, & Boring, 2010; Hart, 2006). It has been shown to be a reliable measure of differences in workloads between tasks in many different conditions (Hart, 2006).

.

4.2.2.2.4.2 Secondary Task Performance

The most common objective measure of workload is performance on a secondary task. Secondary tasks are tasks that people must perform in addition to the primary task. The reason secondary tasks are used is that they allow measurement of the spare attentional capacity that remains from the primary task. As workload on the primary task increases, secondary task performance degrades (Miller, 2001), particularly if the secondary task requires the same cognitive resources as the primary task (Hwang et al., 2008; Miller, 2001; Le Blanc et al., 2010). There are a wide variety of possible secondary tasks (including but not limited to tracking, monitoring, detection, choice reaction time, mental mathematics, classification, and memory scanning tasks Common secondary tasks used in many studies in the nuclear power domain use some variety of a target or signal detection task (e.g., Jou et al., 2009; Lin, Yenn, & Yang, 2010).

4.2.2.3 Timing Metrics

In simulator studies many researchers have used metrics such as time-to-initiate-a-task (Hallbert, Sebok, & Morrisseau, 2000; Converse, 1995) and time-to-complete-a-scenario (Huang & Hwang, 2009; Lois et al., 2009; Xu et al., 2008; Lee, Hwang & Wang, 2005; Converse, 1995). These metrics can be used to evaluate efficiency (i.e., does this help the operator do their job faster?) and safety (i.e., did the task get completed in the required amount of time?).

4.2.2.4 Team Performance Metrics

Given the size and complexity of a commercial NPP, teamwork is a necessity for safe and productive operations. INPO recognizes this and has requirements for training and licensing that focus not only on ensuring the operators have the requisite technical knowledge to operate a plant, but also on an understanding of how to apply the fundamentals of good teamwork to their day-to-day operations. Reports INPO 88-003 (1988), INPO SOER 96-1 (1996), and INPO ACAD 10-001 (2010) provide details on INPO's guidance for the industry, which emphasizes the importance of team performance concepts.

These include communication, leadership, and team dynamics such as constructive conflict management, performing peer-checks, and how to mitigate groupthink (Janis, 1982).

Team performance has both social psychological and cognitive components. While social psychological aspects of teamwork are important (e.g., leadership style, having a positive attitude about working together as a team, actively caring for and trusting each other, holding one another accountable, and not fearing constructive conflict), this research project is focused on the cognitive components of effective teamwork. Even with this novel focus, there are nevertheless a number of research studies that have examined the cognitive aspects of team performance in NPP crews. For example, Roth, Mumaw, & Lewis (1994) performed in-the-field observations of crews in the control room and identified three types of cognitively demanding situations where specific types of crew interaction appeared to contribute positively to successful crew performance. They are: 1) when operators need to pursue multiple objectives, 2) when the situation assessment requires integration of information that is distributed across crewmembers, and 3) when crews have to evaluate the appropriateness of a procedure path and/or decide whether to take actions not explicitly specified in the procedures. Similarly, Park, Jung, and Yang (2012) studied how communication among NPP crews during simulated emergencies affected crew cognition and subsequent performance, and found that: 1) a tightly coupled communication structure (which is an indicator of good team cohesion), 2) increasing the amount/density of communication to increase team situation awareness (e.g., crew members speaking up when observing changes in the system state), and increasing the thoroughness of communication to make shared understanding more explicit (e.g., greater adherence to three-way communication practices) improved overall crew cognition and performance during these scenarios.

Hallbert, Sebok, and Morisseau (2000) studied how the type of simulated control room (i.e., a control room with a conventional HSI vs. a control room with an advanced HSI) and staffing levels affected situation awareness, workload, team interactions, rated crew performance, and objective performance. Overall, they found that the control room with the advanced interface generally improved team and overall system performance, but not necessarily their situation awareness (see Sebok, 2000). They also found that the effect of staffing levels depended on the type of control room. Specifically, the design of a conventional control room tends to follow a single-sensor-single-indicator design philosophy. Because of this, it requires the human operator to do the heavy cognitive processing of synthesizing the single indicator data points into meaningful information. On the other hand, the advanced control room design philosophy uses automated systems programmed with advanced algorithms to do more of the information processing and synthesis of data into meaningful information. Hallbert, Sebok, and Morriseau (2000) found that the change in “who” was responsible for translating data into meaningful information also affected the size of crew needed to perform well. Larger crews in conventional settings were needed to help with the burden of processing of data into information imposed by the HSI. The ‘additional’ crewmember was not needed in advanced control rooms. However, Hallbert, Sebok, and Morriseau (2000) also pointed out that workload was higher for crews in the advanced control rooms, that this was exacerbated when the crew size was smaller, and that while this did not affect their performance during the study, higher workload levels over prolonged periods may adversely affect performance.

Chung, Yoon, and Min (2009) studied how communication protocols vary depending on if the crew is operating in a conventional main control room (CMCR) or an advanced main control room (AMCR). Like Hallbert, Sebok, and Morriseau (2000), they noted that the division of labor between humans and automated systems differs between CMCRs and AMCRs. Based on this, they deconstructed the communications between crewmembers in both the CMCR and AMCR and identified how communication errors would subsequently affect crewmembers’ cognitive processes (e.g., situation awareness), their actions, and the overall system performance. They found that the change in the roles and responsibilities of the crewmembers when in an AMCR not only affected the nature and content and of their communications, but also the types of communication errors they were likely to commit, which subsequently affected their cognition and the likelihood of taking incorrect actions that could challenge plant safety or adversely affect productivity.

Mumaw, Roth, Vicente, and Burns (2000) also performed in-the-field observations of crews and found that one of the primary cognitive tasks of the crews—monitoring plant conditions—is not done in a quiet setting where the onset of abnormal indications is always readily apparent. Rather, abnormal indications are often presented to the operators in a noisy environment with many additional indications that can potentially distract the operator, but more often just make it difficult to discern what information is important. In a follow up study, Vicente, Roth, and Mumaw (2001) noted that, given the existing technology and available indications in the control room, operators often engaged in behaviors that would do one of three things:

1. Enhance information extraction by increasing the salience of important indicators and reducing the background “noise”.
2. Create new information.
3. Offload some of the cognitive processing onto the interface (e.g., creating external aids and reminders for monitoring).

All of the research on team performance cited in this section can be directly related to the technological benefits identified in Table 2. In the case of Vicente, Roth, & Mumaw (2001), the fact that operators found workarounds to create the indications and/or assistance they wanted is evidence that there is a gap in the usefulness of existing control room technologies and indications. These gaps can and should be filled with new advanced control room technologies. Table 2 also makes it apparent that these advanced control room technologies can help address the communication issues identified by Roth, Mumaw, and Lewis (1994), and facilitate the best practices for communication identified by Park, Jung, and Yang (2012), thereby improving overall crew and system performance. And finally, the results from Hallbert, Sebok, and Morriseau (2000) and Chung, Yoon, and Min (2009) provide preliminary evidence that these new technologies, which leverage the design philosophy of AMCRs, combined with appropriate changes in the conduct of operations and new communication protocols that minimize communication errors, are well positioned to demonstrate the benefits this plan is presenting.

Having said this, it is also important to note that many previous studies used an experimental design and an approach to measuring team performance that are different than the planned approach for this benefits project. Most of the studies looking at crew performance involved human factors experts making qualitative, in-the-field observations of crews. Some studies were run in a simulator, but focused on communication patterns and did not include traditional human factors measures of crew performance. Though not discussed in this section, it is worth noting that EPRI’s technical report 1010042 (2005) discussed the importance of teamwork, but focuses on describing how team dynamics can change when new control room technologies are installed, and provides guidance on how to evaluate the design of HSIs that facilitate teamwork. Only two studies previously cited include quantifiable measures of team performance. Hallbert, Sebok, and Morriseau (2000) and HUPESS (Ha & Seong, 2009) both use behaviorally anchored rating scales (BARS), which includes rating scales for observers to assess various dimensions of team performance (e.g., effectiveness of communication, coordination of effort, task focus, and team morale). Thus, it appears that developing a diversified approach to measuring team performance quantitatively in control room simulator studies has not received a lot of attention. However, human factors literature outside the nuclear domain provides other approaches to measuring team performance besides BARS. For example, as a way to identify and mitigate the frequently adverse effects of automation on teamwork, Roessingh and Zon (2004) developed a suite of measures to assess how skills, knowledge, and pro-teamwork attitudes are affected when automation in process control (e.g., air traffic management) is used.

Additionally, Endsley and Jones (1997) extended Endsley’s original situation awareness construct to include team and shared situation awareness. They discuss how teams must have both team situation awareness (e.g., the degree to which all team members have situation awareness that is commensurate with their roles and tasks), and shared situation awareness (e.g., the extent to which team members have the same situation awareness on information that is shared and should be known). Endsley, Bolte, and Jones (2003) further describe how team and share situation awareness can be measured by simply modifying how SAGAT is administered and extending how the results are analyzed. To use SAGAT to measure team and shared situation awareness, it would be given to all team members during the scenario, but before doing so, its questions must be customized for each team member such that it assesses the team member's situation awareness that is commensurate with their role. As such, team members with similar roles would get similar versions of the SAGAT, and members with different roles would get different versions. When administered, team members would complete the SAGAT without conferring with another team member to compare answers. Then, as Endsley, Bolte, and Jones (2003) write, "Team SA can be assessed by evaluating the SA of each team member to determine any weaknesses or SA problems that could undermine team performance. Shared SA can be examined by comparing the team members' responses to the SAGAT queries *on those questions they have in common* (determined by the degree of overlap in their SA requirements)" (pg. 242, italics from the original text). Jones and Endsley (2002b) provide additional detail how these modifications to SAGAT can be used to effectively measure team and shared situation awareness. Additionally, Childs, Ross, and Ross (2000) developed an alternative approach to assess team situation awareness that involves deconstructing team interactions. Overall, there is other research on measuring team performance besides using BARS that should be examined more closely for applicability and usefulness to this benefits research project.

5. Path Forward

This report has introduced candidate control room technology, discussed the potential benefits of those technologies, and described a general approach to conducting research to evaluate those benefits. This section describes the approach to establish an industry collaboration partner, and the approach to each phase of the research.

5.1 Evaluation Methodology

This section describes the general approach to the series of evaluation studies that will be conducted to demonstrate benefits of control room technologies.

5.1.1 Industry Partner

The research team will work closely with a utility partner to conduct this research. The team will identify a host NPP and/or utility to collaborate. The partner will host baseline studies in their training simulator (if possible) and provide operating crew to participate in evaluation studies in the HSSL as outlined in Section 4.1. They will also provide access to process experts to advise on scenario design and the simulation model to use in the HSSL.

5.1.2 Scenario Development

Scenarios will be developed specifically to test the benefits of the candidate technologies selected. The scenarios used will be tailored specifically to the technologies used, and the proposed benefits of those technologies. See Appendix A for a summary of the relationship between the candidate technologies, the performance metrics, and the approach to scenario design. In general, the researchers anticipate developing scenarios to evaluate:

- Normal, routine activities
- Unanticipated activities
- Anticipated transients
- Anticipated design basis accidents
- Beyond design basis accidents (contingent on expected upgrades to HSSL)

5.1.3 Phase One

The first phase of the research will be a pilot test of the evaluation methodology that will be conducted in the HSSL. Several of the near-term candidate technologies will be installed in the HSSL and integrated into an existing simulation model. The researchers will then evaluate performance using the metrics described in Section 4.2 with and without the candidate technologies. Participant for the first phase of research will be licensed operators (if available) or an ad hoc sample of retired operators and/or nuclear engineering students. The purpose of the pilot test is to test and refine the experimental methodology in preparation for the next phase of the research. This phase will occur in Fiscal Year (FY) 2015.

5.1.4 Phase Two

The second phase of the research will be a baseline measure of performance on the scenarios that are developed to test the technologies. This baseline will be conducted in the host plant's training simulator using the host plant's operating crews. The researchers will measure several operating crew's performance, SA workload, and team performance for each of the scenarios. This phase will be conducted in FY 2016.

5.1.5 Phase Three

The third phase of the research will evaluate the candidate technologies in the HSSL. The host plant's simulation model will be installed in the HSSL, and the candidate technologies will be integrated into the simulations. The researchers will invite operating crews from the host utilities to conduct the scenarios in the HSSL with and without the candidate technologies. If resources allow, the researchers will compare performance, SA, workload, and the other metrics described in Section 4.2 using the technologies in the HSSL to baseline performance without the technologies in the HSSL and performance in the host plant's training simulator. This phase will be iterative; each successive study will build on the results of the previous studies. The near-term technologies will be tested first and, as they become available, the farther term technologies will be integrated and tested. This phase will commence in FY 2016 (as resources and availability of industry partners allows) and continue as new technologies are made available for study purposes.

6. References

Bedny, G. Z., & Meister, D. (1999). Theory of activity and situation awareness. International Journal of Cognitive Ergonomics, 3(1), 63-72.

Boring, R., Agarwal, V., Fitzgerald, K., Hugo, J., and Hallbert, B. (2013). Digital Full-Scope Simulation of a Conventional Nuclear Power Plant Control Room, Phase 2: Installation of a Reconfigurable Simulator to Support Nuclear Plant Sustainability, INL/EXT-13-28432. Idaho Falls: Idaho National Laboratory.

Boring, R., and Joe, J. (2014). Baseline Human Factors and Ergonomics in Support of Control Room Modernization at Nuclear Power Plants, INL/EXT-33223. Idaho Falls: Idaho National Laboratory.

Boring, R., Lew, R., Ulrich, T., & Joe, J. (2014). Operator Performance Metrics for Control Room *Modernization: A Practical Guide for Early Design Evaluation, INL/EXT-14-31511*. Idaho Falls: Idaho National Laboratory.

Boring, R.L. (2011). The use of simulators in human factors studies within the nuclear industry. In A.B. Skjerve and A. Bye (eds.), Simulator-Based Human Factors Studies Across 25 Years (pp. 3-17). London: Springer-Verlag.

Boring, R.L., & Joe, J.C. (2014). *Baseline Human Factors and Ergonomics in Support of Control Room Modernization at Nuclear Power Plants, INL/EXT-14-33223*. Idaho Falls: Idaho National Laboratory.

Boring, R.L., Agarwal, V., Joe, J.C., and Persensky, J.J. (2012). Digital Full-Scope Mockup of a Conventional Nuclear Power Plant Control Room, Phase 1: Installation of a Utility Simulator at the Idaho National Laboratory, INL/EXT-12-26367. Idaho Falls: Idaho National Laboratory.

Boring, R.L., Joe, J.C., Ulrich, T.A., & Lew, R.T. (2014). Early-stage design and evaluation for nuclear power plant control room upgrades. Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting, 1909-1913.

Braseth, A., Nihlwing, C., Svengren, H., Veland, Ø., Hurlen, L., Kvalem, J. (2009). Lessons learned from Halden project research on human systems interfaces. Nuclear engineering and technology, vol.41 no.3.

Braun, C., Grimes, J., & Shaver, E., (2011). A Human Factors Perspective on Alarm System Research & Development 2000 to 201. INL/EXT-*11-21498.* Idaho Falls: Idaho National Laboratory

Carvalho, P. V., dos Santos, I. L., Gomes, J. O., Borges, M. R., & Guerlain, S. (2008). Human factors approach for evaluation and redesign of human–system interfaces of a nuclear power plant simulator. *Displays*, *29*(3), 273-284

Chung, Y. H., Yoon, W. C., & Min, D. (2009). A model-based framework for the analysis of team communication in nuclear power plants. Reliability Engineering & System Safety, 94(6), 1030-1040.

Chung, Y.H., Daihwan Min; Kim, B.R. (2002). Observations on emergency operations using computerized procedure system. *Proceedings of the 2002 IEEE 7th Conference on Human Factors and Power Plants*, 4-61- 4-65.

Converse, S. (1995). Evaluation of the Computerized Procedure Manual II (COPMA II) (NUREG/CR-6398). U.S. Nuclear Regulatory Commission, Washington

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32-64. doi:10.1518/001872095779049543

Endsley, M. R. (2000). Direct measurement of situation awareness: Validity and use of SAGAT. In M. R. Endsley & D. J. Garland (Eds.), *Situation Awareness Analysis and Measurement* (pp. 130-148) Mahwah, NJ: Lawrence Erlbaum Associates.

EPRI (2010), "Computerized Procedure Systems: Guidance on Design, Implementation and Use of Computerized Procedure Systems, Associated Automation and Soft Controls," EPRI 1015313, Electric Power Research Institute, Palo Alto, CA.

EPRI (2005). *Human Factors Guidance for Control Room and Digital Human-System Interface Design and Modification: Guidelines for Planning, Specification, Design, Licensing, Implementation, Training, Operation, and Maintenance, EPRI TR-1010042*. Palo Alto, CA: EPRI.

Fink, R., Killian, C., Hanes, L., & Naser, J. (2009). Guidelines for the design and implementation of computerized procedures. *Nuclear News. VOL* 52 (3), 85-88, 90.

Ha, J. S., & Seong, P. H. (2009). HUPESS: Human performance evaluation support system. In P. Seong (Ed.), *Reliability and Risk Issues in Large Scale Safety-critical Digital Control Systems*, 197-229. Springer, London, UK.

Ha, J., Seong, P., Lee, M., & Hong, J. (2007). Development of Human Performance Measures for Human Factors Validation in the Advanced MCR of APR-1400. *IEEE Transactions on Nuclear Science, Vol. 54, No. 6.*

Hallbert, B. P. & Sebok, A. & Morrisseau, D. & U.S. Nuclear Regulatory Commission. Office of Nuclear Regulatory Research. & OECD Halden Reactor Project. (2000). A study of control room staffing levels for advanced reactors [microform] / prepared by B.P. Hallbert, A. Sebok, D. Morrisseau Office of Nuclear Regulatory Research, U.S. Nuclear Regulatory Commission, Washington, DC..

Hallbert, B., Morgan, T., Hugo, J., Oxstrand, J., & Persensky, J. (2014). Formalized Approach for the Collection of HRA Data from Nuclear Power Plant Simulators, NUREG/CR-7163. U.S. Nuclear Regulatory Commission, Washington, D.C.

Hallbert, B., Sebok, A., & Morisseau, D. (2000). A study of control room staffing levels for advanced reactors, NUREG/IA-0137. U.S. Nuclear Regulatory Commission, Washington, D.C.

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 50, No. 9, pp. 904-908). Sage Publications.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*,*52*, 139-183.

Hogg, D. N., Follesø, K., Strand-Volden, F., & Torralba, B. (1995). Development of a situation awareness measure to evaluate advanced alarms systems in nuclear power plant control rooms. *Ergonomics, 38*(11), 2394-2413.

Huang, F. H., Hwang, S. L., Yenn, T. C., Yu, Y. C., Hsu, C. C., & Huang, H. W. (2006). Evaluation and comparison of alarm reset modes in advanced control room of nuclear power plants. *Safety science*, *44*(10), 935-94

Huang, F.-H., & Hwang, S-L. (2009). Experimental studies of computerized procedures and team size in nuclear power plant operations. Nuclear Engineering and Design. 239, 373–380.

Hugo, J., Boring, R., Hanes, L., & Thomas, K. (2013). *A Reference Plan for Control Room Modernization: Planning and Analysis Phase, INL/EXT-13-30109.* Idaho Falls: Idaho National Laboratory..

Hwang, S. L., Yau, Y. J., Lin, Y. T., Chen, J. H., Huang, T. H., Yenn, T. C., & Hsu, C. C. (2008). Predicting work performance in nuclear power plants. *Safety science*, *46*(7), 1115-1124.

IEEE (2009). Human Factors Guide for Applications of Computerized Operating Procedure Systems at Nuclear Power Generating Stations and other Nuclear Facilities (Draft Guideline 1786, 9/30/09 version). New York, NY: Institute of Electrical and Electronics Engineers.

INPO (1988). Guideline for Teamwork and Diagnostic Skill Development (INPO 88-003). Atlanta, GA: Institute for Nuclear Power Operations.

INPO (1996). Control Room Supervision, Operational Decision-Making, and Teamwork (INPO SOER 96-1). Atlanta, GA. Institute for Nuclear Power Operations.

INPO (2010). Guideline for Training and Qualification of Licensed Operators (INPO 10-001). Atlanta, GA. Institute for Nuclear Power Operations.

International ElectroTechnical Commission (IEC) 62241. (2010). Nuclear Power Plants Control Rooms Alarm Function and Presentation, the IEC National Committee, in Compliance with the ISO/ISO directives, Geneva, Switzerland

Jang, I., Park, J., & Seong, P. (2012). An empirical study on the relationships between functional performance measure and task performance measure in NPP MCR. Annals of Nuclear Energy, 42, 96-103.

Janis, I. L. (1982). Groupthink: Psychological Studies of Policy Decisions and Fiascoes. Wilmington, MA: Houghton Mifflin.

Jeffroy, F., & Charron, S. (1997, June). From safety assessment to research in the domain of human factors: the case of operation with computerised procedures. In *Human Factors and Power Plants, 1997. Global Perspectives of Human Factors in Power Generation., Proceedings of the 1997 IEEE Sixth Conference on* (pp. 13-5). IEEE.

Joe, J.C., Boring, R.L., & Persensky, J.J. (2012). Commercial utility perspectives on nuclear power plant control room modernization. *8th International Topical Meeting on Nuclear Power Plant Instrumentation, Control, and Human-Machine Interface Technologies (NPIC&HMIT),* 2039-2046.

Jou, Y.-T., Yenn, T.-C., Lin, C. J., Yang, C.-W., & Chiang, C.-C. (2009). Evaluation of operators' mental workload of human-system interface automation in the advanced nuclear power plants. *Nuclear Engineering and Design, 239,* 2537-2542.

Klein, G. A., Moon, B. M., & Hoffman, R. R. (2006). Making sense of sensemaking. Part 2: A macrocognitive model. IEEE Intelligent Systems, 21(5), 88-92. doi:10.1109/mis.2006.100.

Laarni, J., Koskinen, H., Salo, L., Norros, L., Braseth, A. O., & Nurmilaukas, V. (2009). Evaluation of the Fortum IRD pilot. In *Proceedings of the Sixth American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies NPIC&HMIT.*

Le Blanc, K., Gertman; D., & Boring, R. (2010). Review of methods related to assessing human performance in nuclear power plant control room simulations. Paper presented at the *Seventh American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies* (NPIC&HMIT), Las Vegas, NV, November 7-11, 2010. LaGrange Park, IL: American Nuclear Society.

Le Blanc, K., Oxstrand, J., Waicosky, T. (2012). Model of procedure usage – results from a qualitative study to inform design of computer-based procedures. In *Proceedings of the Seventh American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies NPIC&HMIT 2012*, San Diego, California

Lee, Y., Hwang, S., & Wang, E. (2004). Reducing cognitive workload of a computer based procedure system. *International Journal of Human-Computer Studies 63, p587-606.*

Lin, C. J., Yenn, T.-C., & Yang, C.-W. (2010a). Automation design in advanced control rooms of the modernized nuclear power plants. *Safety Science, 48*, 63-71

Lois, E., Dang, V.N., Forester, J., Broberg, H., Massaiu, S., Hildebrandt, M., Braarud, P.Ø., Parry, G., Julius, J., Boring, R., Männistö, I., &Bye, A. (2009). International HRA empirical study—Phase 1 report, Description of overall approach and pilot phase results from comparing HRA methods to simulator performance data, NUREG/IA-0216, Vol. 1. US Nuclear Regulatory Commission,Washington, DC.

Manzey, D., Reichenbach, J., & Onnasch, L. (2008). Performance Consequences of Automated Aids in Supervisory Control: The Impact of Function Allocation. In Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting. Santa Monica, CA: Human Factors and Ergonomics Society.

Meister, D. (2004). Conceptual Foundations of Human Factors Measurement. Mahwah: Lawrence Erlbaum Associates.

Miller, S. (2001). *Literature Review: Workload Measures* (Document ID: N01-006). Iowa City, IA: University of Iowa, National Advanced Driving Simulator.

Mumaw, R. J., Roth, E. M., Vicente, K. J., & Burns, C. M. (2000). There is more to monitoring a nuclear power plant than meets the eye. Human Factors: The Journal of the Human Factors and Ergonomics Society, 42(1), 36-55.

NEI (2000). Guidelines for 10 CFR 50.59 Evaluations (NEI 96-07). Washington D.C., Nuclear Energy Institute.

Noyes, M. V., & Sheridan, T. B. (1984). A novel predictor for telemanipulation through a time delay. In Proceedings of Annual Conference on Manual Control. NASA Ames Research Center.

O'Hara, J., Higgins, J. C., Stubler, W. F., and Kramer, J. (2000). NUREG/CR-6634 Computer-Based Procedure Systems: Technical Basis and Human Factors Review Guidance, US Nuclear Regulatory Commission, Washington, DC.

O'Hara, J. M., Brown, W. S., Hallbert, B., Skraning, G., Persensky, J. J., & Wachtel, J. (2000). The effects of alarm display, processing, and availability on crew performance (No. BNL-NUREG-52600). BROOKHAVEN NATIONAL LAB UPTON NY.

Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2013). Human Performance Consequences of Stages and Levels of Automation An Integrated Meta-Analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 0018720813501549Onnasch et al., 2013

Oxstrand, J.H., & Le Blanc, K.L. (2012). Computer-Based Procedures for Field Workers in Nuclear Power Plants: Development of a Model of Procedure Usage and Identification of Requirements, INL/EXT-12-25671. Idaho Falls: Idaho National Laboratory.

Park, J., Jung, W., & Yang, J. E. (2012). Investigating the effect of communication characteristics on crew performance under the simulated emergency condition of nuclear power plants. Reliability Engineering & System Safety, 101, 1-13.

Persensky, J., Boring, R., Le Blanc, K., Hugo, J., Gertman, D., Shaver, E., Braun, C., Oxstrand, J., and Richards, R. (2010). *Alarm System Research Plan: Milestone Deliverable in Support of the U.S. Department of Energy's Light Water Reactor Sustainability Project, INL/EXT-10-19888.* Idaho Falls, ID: Idaho National Laboratory.

Roth, E. M. and O'Hara, J. (2002) *Integrating Digital and Conventional Human System Interfaces: Lessons Learned from a Control Room Modernization Program*. Washington, D. C.: U. S. Nuclear Regulatory Commission. (NUREG/CR-6749 also BNL-NUREG-52638)

Roth, E. M., Mumaw, R. J., & Lewis, P. M. (1994). An empirical investigation of operator performance in cognitively demanding simulated emergencies, (NUREG/CR-6208). Washington, DC, U.S. Nuclear Regulatory Commission.

Rouse, W. B., & Boff, K. R. (2003). Cost–Benefit Analysis for Human Systems Integration. In H. Booher (Ed.) *Handbook of Human Systems Integration*, 631-657, John Wiley & Sons, Inc., USA.

Sebok, A. (2000), 'Team Performance in Process Control: Influence of Interface Design and Staffing Levels', *Ergonomics*, vol. 43 (8), pp. 1210-1236.

Skraaning, G. (2004). Experimental Control Versus Realism: Methodological Solutions for Simulator Studies in Complex Operating Environments, HWR-361, Halden Reactor Project, Halden, Norway.

Smith, A. & Jamieson, G. (2012). Level of Automation Effects on Situation Awareness and Functional Specificity in Automation Reliance. *In Proceedings of Human Factors Society and Ergonomics Society, 56th Annual Meeting.* Santa Monica, CA: Human Factors and Ergonomics Society.

Smith, K., & Hancock, P. A. (1995). Situation awareness is adaptive, externally directed consciousness. Human Factors, 37(1), 137-148. doi:10.1518/001872095779049444.

Stanton, N.A., Salmon, P.M., Rafferty, L.A., Walker, G.H., Baber, C., & Jenkins, D.P. (2013). Human Factors Methods: A Practical Guide for Engineering and Design, 2nd Edition. Burlington, VT: Ashgate Publishing Company.

Stubler, W.F, O'Hara, J.M, (1996) Group-view display system: Functional characteristics and review criteria. NUREG 0-700, Washington DC: U.S. Nuclear Regulatory Commission

Taylor, R. M. (1990). Situational Awareness Rating Technique(SART): The development of a tool for aircrew systems design. *AGARD, Situational Awareness in Aerospace Operations 17 p(SEE N 90-28972 23-53)*.
Thomas, K., Boring, R., Lew, R., Ulrich, T., & Villim, R. (2013). A Computerized Operator Support System Prototype, INL/EXT-13-29651. Idaho Falls: Idaho National Laboratory.
U.S. Nuclear Regulatory Commission. (2002). Human-System Interface Design Review Guidelines, Rev. 2. Washington, DC: U.S. Nuclear Regulatory Commission.
U.S. Nuclear Regulatory Commission. (2012). Human Factors Engineering Program Review Model, NUREG-0711, Rev. 3. Washington, DC: U.S. Nuclear Regulatory Commission.
Ulrich, T., Boring, R., & Lew, R. (2014). *Human Factors Engineering Design Phase Report for Control Room Modernization, INL/EXT-14-33221.* Idaho Falls: Idaho National Laboratory.
Ulrich, T., Boring, R., Phoenix, W., DeHority, E., Whiting, T., Morrell, J., and Backstrom, R. (2012). Applying Human Factors Evaluation and Design Guidance to a Nuclear Power Plant Digital Control System, INL/EXT-12-26797. Idaho Falls: Idaho National Laboratory.
van Breda, L. (1999). Anticipatory behaviour in supervisory control. Delft, Netherlands: Delft University Press.
Vicente, K. J., Roth, E. M., & Mumaw, R. J. (2001). How do operators monitor a complex, dynamic work domain? The impact of control room technology. International Journal of Human-Computer Studies, 54(6), 831-856.
Wickens, C. & Hollands, J. (2000). Engineering Psychology and Human Performance (3rd ed.). Upper Saddle River, NJ: Prentice Hall
Wickens, C. D. (2003). Aviation displays. In P. Tsang & M. Vidulich (Eds.), Principles and practices of aviation psychology (pp. 147–199). Mahwah, NJ: Lawrence Erlbaum.
Wickens, C., Li, H., Santamaria, A., Sebok, A., & Sarter, N. (2010). Stages and Levels of Automation: An Integrated Meta-analysis. In *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting*, Santa Monica: CA: HFES.
Wright, M. C., & Kaber, D. B. (2005). Effects of automation of information processing functions on teamwork. Human Factors, 47(1), 50-66.
Xu, S., Song, F., Li, Z.Z., Zhao, Q.Y., Luo, W., He, X.H., and Salvendy, G. (2008). An ergonomics study of computerized emergency operating procedures: presentation style, task complexity, and training level, Reliability Engineering and System Safety 93 (10), 1500–1511.
Yin, S., Wickens, C. D., Helander, M., & Laberge, J. C. (2014). Predictive Displays for a Process-Control Schematic Interface. Human Factors: The Journal of the Human Factors and Ergonomics Society, 0018720814542104.

Appendix A Table of Benefits and Approach to Performance Measurement

Potential Benefit	Technology	Possible Measures	Objective Metrics	Subjective Metrics	Scenario Design Considerations
Reduced mental workload during upset conditions	CBPs, Alarm System	• Mental Workload	• Performance on a secondary task • Eye Blink Data • Other Physiological Measures	NASA TLX	Complex, novel abnormal scenarios
Reduced Physical workload during upset conditions	TBDs, LSDs	• Physical Workload	• Steps (Pedometer) • Physical control actions	Physical workload sub-scale of NASA TLX	Complex, novel, abnormal scenarios
Reduced/balanced mental workload during normal operations	CBPs, Alarm System, LSDs, TBDs	• Mental Workload	• Performance on a secondary task • Eye Blink Data • Other Physiological Measures	NASA TLX	Normal Operating Scenarios
Increased Efficiency E.g., Reduced need for peer-checks Reduced shift turnover time (reduced shift briefs time etc.)	CBPs, LSDs, TBDs	• Time-complete various tasks • Performance based on crew size	• Completion time • Response time • Performance based on crew size	• Expert observation (e.g., SAT/UNSAT scores)	Normal operating scenarios
Increased SA/ Increased Understanding of Plant Status	LSD, CBPs, Alarm system	• Situation Awareness	• SACRI • Eye tracking	• Subjective SA assessment (e.g. SART)	A wide variety of scenarios

Potential Benefit	Technology	Possible Measures	Objective Metrics	Subjective Metrics	Scenario Design Considerations
Enhanced Teamwork	LSD, CBPs	• Team SA • Team Performance • Communication performance	• Communication analysis • Crew-System Performance	• Behaviorally Anchored Ratings scales • Expert observation (e.g., SAT/UNSAT scores)	Scenarios that require team coordination
Understanding of plant status at-a-glace	LSD	• At-a-glance SA	• SACRI-like measure for at-glance status display	• SA Subjective SA assessment (E.g. SART)	Don't need scenarios, per se, could simply utilize technologies (e.g., large screen overview displays)and have operators assess conditions
Increased ability to adapt to changing conditions, especially unanticipated conditions	LSD Alarm, System	• Plant Performance • Situation Awareness • Fault management	• Deviation in Plant Parameters • SACRI • Eye tracking	• Expert observation (SAT/UNSAT) • Subjective SA assessment	Novel scenarios in which procedural guidance may not apply
Reduced Training time	All	• Training time • Learnability	• Time-to proficiency • Performance during training	• Learnability survey	Training scenarios that demonstrate how to use technology.
Quicker identification of transients/abnormal conditions	Alarm System, LSD	•	• Time-to-initiate procedure • Time to diagnose	• N/A	Scenarios with transients/abnormal conditions
Ability to operate with broader safety margins	CBPs, Alarm System, LSDs, TBDs	• Safety Margin	• Compare margin of safety based on observed parameters and tech specs	• N/A	Any Scenario

Potential Benefit	**Technology**	**Possible Measures**	**Objective Metrics**	**Subjective Metrics**	**Scenario Design Considerations**
Reduced stress on plant equipment	Alarm System, LSD	• Stress on plant equipment	• Compare Tech Specs to simulator logged values	• N/A	This can be measured using any scenario, although it will be more sensitive with scenarios that are likely to case stress on plant equipment.
Enhanced Procedure Performance	CBPs, TBDs	• Time to complete Procedure • Procedure Execution accuracy • Plant performance during Procedure execution • Crew and Individual Performance	• Time to complete Procedure • Discrepancy Sores	• Expert observation (SAT/UNSAT) • Individual rating of performance	Scenarios that require procedures

TBD = Task Based Display, LSD = Large Screen Display, CBP = Computer-Based Procedures