INL/EXT-17-43188

Light Water Reactor Sustainability Program

Analog, Digital, or Enhanced Human-System Interfaces? Results of an Operator-in-the-Loop Study on Main Control Room Modernization for a Nuclear Power Plant

September 2017

U.S. Department of Energy

Office of Nuclear Energy

INL/EXT-17-43188

Analog, Digital, or Enhanced Human-System Interfaces? Results of an Operator-in-the-Loop Study on Main Control Room Modernization for a Nuclear Power Plant

Ronald Boring[1], Thomas Ulrich[1], Roger Lew[2], Casey Kovesdi[1], Brandon Rice[1], Christopher Poresky[3], Zach Spielman[1], and Kateryna Savchenko[1]

[1]Idaho National Laboratory
[2]University of Idaho
[3]University of California, Berkeley

September 2017

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

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

(This page intentionally left blank)

ABSTRACT

The second Operator Study of System Overviews (OSSO-2) was conducted with a three-person reactor operator crew in the Human Systems Simulation Laboratory (HSSL) at Idaho National Laboratory (INL) in August, 2017. The study supported control room modernization at a nuclear power plant and featured a benchmark comparison of three variants of a turbine control system (TCS): the existing analog TCS, a proposed standalone digital TCS with two displays, and the digital TCS with the addition of a third display consisting of a system overview screen. TCS prototypes were developed at INL to allow evaluation of operator performance and preferences during realistic turbine scenarios in the full-scope simulator. The study revealed that completion of turbine startup was several minutes faster with the digital TCS variants than with the conventional analog TCS. Eye tracking scan paths were more widely distributed in the overview vs. standalone TCS condition, suggesting the overview screen was cueing reactor operators to verify values across the boards. There was no significant difference in workload or situation awareness across the three interfaces. Reviewing key plant parameters showed smoother transitions during load following for the digital vs. analog TCS. Despite some performance advantages for the digital TCS variants, operators preferred the existing analog TCS. Open-ended responses suggested this finding may be more an artifact of familiarity than a reflection of dissatisfaction with the new TCS. OSSO-2 provides compelling evidence that the new digital TCS was used successfully by the operators without extensive training of rewriting of the operating procedures, suggesting high usability for the digital TCS design. Further advantages were realized through the addition of the system overview screen to provide crews with at-a-glance indicators of key turbine parameters.

(This page intentionally left blank)

ACKNOWLEDGMENTS

This report was made possible through funding by the U.S. Department of Energy (DOE) Light Water Reactor Sustainability (LWRS) Program. We are grateful to the crews, instructors, simulator staff, and engineering staff from the participating nuclear utility, without whom this study would not have been possible. Additionally, we gratefully acknowledge design and technical assistance from staff at Idaho National Laboratory (INL), including Bruce Hallbert, Ken Thomas, and Jacques Hugo. Drs. Wondea Jung and Jinkyun Park from the Korea Atomic Energy Research Institute (KAERI) provided insights into human performance measures used for assessing operator reliability in digital control rooms. We thank them for their helpful inputs to the design of this study.

(This page intentionally left blank)

CONTENTS

FIGURES

TABLES

ACRONYMS

ALARA	As Low As Reasonable Assessment
ANIME	Advanced Nuclear Interface Modeling Environment
AOI	Area of Interest
BOP	Balance of Plant
CRS	Control Room Supervisor
CVCS	Chemical and Volume Control System
DCS	Distributed Control System
DOE	Department of Energy
GONUKE	Guideline for Operational Nuclear Usability and Knowledge Elicitation
HSI	Human-System Interface
HRA	Human Reliability Analysis
HSSL	Human Systems Simulation Laboratory
I&C	Instrumentation and Controls
INL	Idaho National Laboratory
ISV	Integrated System Validation
KAERI	Korea Atomic Energy Research Institute
LWRS	Light Water Reactor Sustainability
NASA	National Aeronautics and Space Administration
NEET	Nuclear Energy Enabling Technology
NPP	Nuclear Power Plant
OAC	Operator at Controls
OSSO	Operator Study of System Overviews
PSF	Performance Shaping Factor
RMS	Root Mean Square
SA	Situation Awareness
SART	Situation Awareness Rating Technique
STA	Shift Technical Advisor
Tavg	Average Temperature
TCS	Turbine Control System
TLX	Task Load Index
U.S.	United States
V&V	Verification and Validation
WSC	Western Services Corporation

1. INTRODUCTION

The authors of this report have previously reported on the need for control room modernization in nuclear power plants (NPPs) in the U.S. (Boring, 2013 and 2014; Boring and Lau, 2016). Idaho National Laboratory (INL), through funding from the U.S. Department of Energy's (DOE) Light Water Reactor Sustainability (LWRS) Program, has developed processes for prototyping upgraded systems and evaluating the performance of those systems (Boring and Joe, 2015; Boring et al., 2015; Boring et al., 2017). Working with simulator vendors, system vendors, and utilities, INL configures plant simulators with digital instrumentation and control (I&C) systems that are then evaluated using operator-in-the-loop studies. The studies to date have served primarily to benchmark and refine upgrade designs, but increasingly the validation and verification (V&V) of the upgraded systems is also important. By establishing a human factors approach to the upgrades, in which operator performance and preferences are considered in the design of the system, it is possible to ensure that new interfaces are as safe and usable as their predecessors.

The main control rooms of NPPs in the U.S. were designed up to a decade prior to installation, and most NPPs in the U.S. in current operation are approaching or have extended their original 40-year operating license. The vintage of the main control rooms belies their reliability. Analog and mechanical I&C have operated reliably for the life of the plant, and a stockpile of replacement parts has ensured the continuous operation of the plants. However, over the lifespan of the plants, the digital revolution occurred, rendering most original equipment manufacturers to abandon their analog I&C in favor of digital technology. There remain few suppliers to make comparable I&C, which over time forces the need for modernization in the main control room. There is no imperative that the technology does not work; the imperative is simply that the technology is no longer available or maintainable.

Modernizing main control rooms is complex, since a variety of safety, operational, engineering, regulatory, and financial considerations come into play. As already noted, the new technology must be at least as safe as the technology it replaces. Installation of the upgraded system must be minimally disruptive to the operation of the plant, because downtime in the main control room will result in extended outages. From an engineering standpoint, fundamentally incompatible analog and digital technologies must be blended, and even small changes to board layouts can have big consequences to the control technology behind the boards. Some changes will require license modification, and grandfathered systems may come under scrutiny to comply with current standards. Moreover, there remains little practical experience conducting major main control room upgrades in the U.S., leaving the regulatory review process somewhat uncharted. Finally, all upgrades must be considered in terms of their cost. Because plants are already operating at peak efficiency and operators already have an impressive record of safety and human reliability, the question of cost without return on investment is a priority. Control room modernization is more than maintenance, because it is a fundamental reworking of the control mechanism and operational concept of the plant. Yet, such changes are inevitably more expensive than maintenance, and the benefits of the switch to digital such as avoiding unnecessary plant downtime—while necessary from an obsolescence standpoint—are not always guaranteed.

The process advocated by INL is the stepwise upgrade, in which one system or subsystem is upgraded during an outage. The window of time to complete such an upgrade is sufficient that it may be successfully deployed during a typical thirty-day outage. Of course, all engineering and design work must be successfully executed prior to the deployment. INL provides human factors and design expertise to ensure that systems are suitable for operator use prior to factory acceptance testing. Prototypes are built to specifications using rapid application development techniques and tested with operators. In this manner, systems may be iterated and refined at the specification phase, resulting in a human factors compliant and operationally tested interface ready for installation.

This report chronicles a recent operator-in-the-loop study focused on the stepwise introduction of a turbine control system (TCS) into an existing multi-unit NPP in the U.S. The plant is not named here, as the goal of this report is to disseminate findings across the nuclear industry rather than offer plant-specific findings. TCS is one of the upgrades being most frequently implemented in NPPs. The technology is well established from non-nuclear applications like fossil fuel generating stations; the turbine is not considered part of the safety critical systems that are designed to protect the reactor and that would require extensive license modifications; it is often possible to achieve additional electricity production and increased plant revenue through upgraded turbine components including the control system; and there is in some cases the opportunity to take advantage of underused features like load following (i.e., adjusting the total electrical output of the plant to complement the output of fluctuating energy sources like renewables), which might strain the capabilities of existing control systems.

The current study involved a single crew carrying out scenarios to test new human-system interfaces (HSIs) during August, 2017. The study was called Operator Study of System Overviews 2 (OSSO-2) and builds on the similar OSSO-1 study conducted with two crews the previous year (Al Rashdan et al., 2017). OSSO-1 and -2 are centered on exploring three HSI variants:

- Existing analog control boards,
- Proposed digital HSIs, and
- Proposed digital HSIs with overview screens.

This study allows clear baselining of performance on existing systems, benchmarking performance of new vendor-proposed systems, and consideration of possible enhancements to the proposed systems. OSSO-1 and -2 considered enhancements particular to system overview screens.[1] Overviews, also called dashboards, provide key information necessary to maintain situation awareness of plant processes. Safety parameter display systems (SPDS; U.S. Nuclear Regulatory Commission, 1981) are early examples of such screens, which were required as part of the industry response to issues identified in the Three Mile Island plant meltdown. The SPDS presents key plant parameters at a glance to the operators so that they can always monitor the overall functioning of the plant. Modern variants of overview screens can be seen in the plant

[1] In this report, the terms *screen* and *display* are used somewhat synonymously. It is acknowledged that a screen commonly refers to the information that is rendered, while a display commonly refers to the physical device upon which information is rendered. In the present use, we are not concerned with the physical device, but rather the content that may be rendered on many different types of physical devices either on the boards or standalone at workstations.

overview displays developed at Halden Reactor Project (Jokstad and Boring, 2015). The question presented in OSSO-1 and -2 is the extent to which advantages of plant overviews and SPDS can also be realized for subsystems of the plant. There remains little definitive guidance on the optimal presentation of such information. Most distributed control system (DCS) screens, such as found in a digital TCS, feature screens with nested information, whereby operators navigate to key information and control screens within a functional hierarchy. The disadvantage of nested information is that some information may be occluded, depending on which screens are active. Many DCS installations feature two physical displays side-by-side, which means different screens can be seen concurrently to minimize information occlusion. Conveniently, the double-display format also provides hardware redundancy in the event of failure of one of the displays.

OSSO-1 and -2 explore the design of system overviews to determine if there are performance advantages afforded by the screens and what the elements of those screens should be. OSSO-1 focused on the elements of the overview screens, while the present study, OSSO-2, centers on measuring operator performance with regard to representative system overviews.

OSSO-1 and -2 featured two different types of systems—TCS and chemical and volume control system (CVCS). The CVCS system overview featured a prognostic fault detection system sponsored by the U.S. DOE’s Nuclear Energy Enabling Technology (NEET) Program. While the TCS portion of the study is centered on addressing immediate needs of nuclear utilities as they modernize their main control rooms, the CVCS prognostic system is forward looking to future technologies not ready for immediate implementation of commercialization. For this reason, the CVCS system overview is documented in a separate report (Ulrich et al., 2017). The present report addresses only the findings from the three HSI variants of the TCS.

(This page intentionally left blank)

2. STUDY DESIGN AND METHOD

2.1 Purpose

The purpose of the operator-in-the-loop study was threefold:

- Determine any issues with the current analog TCS and any features and functions that operators would like to retain.
- Determine the suitability of the vendor-proposed digital TCS upgrade.
- Determine the value of adding system overview screens to the TCS upgrade.

To gather these insights, the full-scope training simulator from the NPP was installed on the glasstop simulator at INL, scenarios were established to evaluate a range of activities associated with the TCS, and functional prototypes of the digital TCS upgrade and overview screen were developed.

2.2 Scenarios

The study centered around operators performing a variety of scenarios. While a number of small fault scenarios were examined relative to the TCS, the emphasis of the study was on benchmarking the three variants of the HSI across two main scenarios:

- *Turbine startup*—This scenario encompasses almost all functions of the TCS, from initial steam flow and warming, to testing and tripping the turbine, to synchronizing the turbine to the grid. Every function and fault possible in the TCS is a variant of activities conducted in startup. In fact, most turbine-centered activities are covered in a common turbine procedure, of which the bulk is startup related.
- *Turbine load following*—While NPPs are designed as baseload systems that may be slow to ramp up or down, with the introduction of renewable energy sources (e.g., solar), it may be desirable to reduce electrical output of the plant in response to the fluctuating availability of such sources (e.g., solar energy is time-of-day dependent). To help address this potential functionality, a scenario was developed in which the electrical output of the plant is reduced by 30% to accommodate the availability of additional electrical sources on the grid. Load following represents a function for which the plant controls were not originally optimized, and it is desirable to see if the task can be made easier for the operators through the advent of the digital HSI.

Note that these scenarios were not counterbalanced in the study. The order strictly followed the evolution of the technology, as depicted in Table 1.

Table 1. Turbine control system scenarios in study.

Order	Scenario	Analog	Digital	Overview
1	Startup	✓	-	-
2	Load Following	✓	-	-
3	Startup	-	✓	-
4	Load Following	-	✓	-
5	Startup	-	✓	✓
6	Load Following	-	✓	✓

Interspersed in these main scenarios were, however, miscellaneous turbine fault small scenarios for the accompanying HSI technology. Additionally, there were scenarios that were not turbine related, e.g., steam generator tube rupture, miscellaneous CVCS faults, and faulty (i.e., spoofed) indicator scenarios. These helped to offset the repetitiveness of the tasking in the scenarios. The benchmark approach used in OSSO is susceptible to possible practice effects of operators performing the same tasks in sequence. However, it was important to evaluate the responses to the respective technologies in sequence. For example, it was necessary to determine reactions to the digital HSI without overview screens before showing the overview screens. Showing the overview screens before the standalone digital TCS HSI might introduce comparisons that would downplay the effectiveness of the digital TCS by itself, if the operators found the overview screens advantageous.

Each of these six scenarios listed above represents extended time frames, requiring over half an hour for the operators to perform the tasks. Just as it is important to have full-scope (i.e., high fidelity) simulations, it is also necessary to have full scenarios for some types of operator evaluations. While some operator performance testing situations may benefit from short scenarios (e.g., Hildebrandt and Fernandes, 2016), there are also clear reasons for undergoing longer scenarios such as employed in this study. Carroll (2000) has articulated advantages of using scenarios in user studies. These include the reflection required by designers to drive the scenario in a realistic manner;[2] the ability to categorize aspects of the scenario as features for evaluation and comparison; and the promotion of communication between stakeholders, designers, and end users.

In addition to Carroll's observations, we offer the following advantages to scenario-based user studies for the purposes of evaluating the TCS upgrade:

- *Performance in context*—Operators may perform differently for short sprint tasks vs. longer duration tasks. When the purpose of studying operator performance is to gauge their overall performance on completing a task involving multiple sets of subtasks, it is desirable to look at the entire changing context of the task and the emergence of operator behaviors throughout that task. For example, workload may change depending on the particular subtask, but it may not be possible to capture true workload if the subtask is

[2] The use of scenarios to create a realistic design piggybacks many efforts to design with a specific user in mind through the use of design personas (Cooper, 2004). The persona and scenario together create a realistic context for designing and evaluating the HSI.

performed in isolation. The conflation of subtasks and procedure steps may produce a different emergent performance than instantaneous subtasks prompted out of process.

- *Operator flow*—Flow is the concept that there is a psychological momentum to performing tasks (Csikszentmihalyi, 1990). To capture the operators' flow through the task, it is important to model the entirety of the task to see where disruptions might occur. While a navigation menu may capture the correct elements according to initial operator screening, the implementation of the menu might prove cumbersome and disruptive to task completion. Flow and disruptions can best be tested by allowing the full scenario and task to play out.
- *Complexity of tasks*—Some tasks are sufficiently complex to tax the operators' ability to complete the task or maintain situation awareness. Complex tasks such as multiple faults or time critical activities like turbine synchronization often involve mentally demanding tasks in parallel or in rapid succession. In simple scenarios, it can be difficult to elicit the high workload and diminished situation awareness that operators experience. While validation of a system does not necessarily involve stress testing the operators, where the task realistically warrants such a complex context, it should be captured in the scenario. Likewise, multiple interleaved tasks merit testing in scenarios.
- *Integration of systems*—The TCS is not operated in isolation (if so, it would be possible to perform studies using a part-task simulator). Rather, changes in demand for steam must be met by corresponding changes in reactor generation. The two systems are interlinked such that it is not truly possible to evaluate operator performance on the TCS without also considering required activities on the reactor side of the house. Where the system under investigation requires interaction with another system, it is warranted to look at scenarios that bring together those systems.
- *Procedure following*—Nuclear power operations feature a high degree of procedural specification in the form of required steps to be taken to complete actions. Scenarios provide an ideal way to evaluate procedure following, especially in the context of a new system for which procedures have not yet been optimized. Procedures formalize a type of task analysis of required activities by the operators. By having operators perform scenarios, it is possible to evaluate the completeness and quality of the procedures, operator procedural adherence, and mismatches between an operators' mental model of a task and the required sequence.

As Nielsen (2014) has pointed out:

> The most effective way of understanding what works and what doesn't in an interface is to watch people use it. This is the essence of usability testing. When the right participants attempt realistic activities, you gain qualitative insights into what is causing users to have trouble. These insights help you determine how to improve the design.

The scenarios of turbine startup and turbine load following on the analog, digital, and overview HSI variants represent the realistic activities the operators need to perform when using the turbine at the plant. These scenarios are the basis for effective operator studies on TCS.

2.3 Simulator Environment

The Human Systems Simulation Laboratory (HSSL; described in detail in Boring et al., 2012 and 2013) represents a reconfigurable, full-scale, and full-scope control room simulator, housed at INL (see Figure 1). The simulator consists of glasstop panels linked together to create a virtual representation of the front panels of the main control room of an NPP. The simulator is driven by the same certified, high fidelity training simulator used at the plants. Whereas at the plants, the simulator model is tied to hardware replicas of the actual control room, complete with analog I&C, the HSSL features virtual versions of the I&C on touchscreen displays. The virtual mimics of the control boards allow operators to see analog instrumentation reflective of the current control room and operate analog controls using gestures on the touchscreen interface. The virtual control boards also allow the introduction of new digital HSIs to reflect upgrades at the plant. Figure 2 provides an example how the boards can be reconfigured from the original layout to include a new HSI. In the figure, the new HSI represents a single-display digital TCS as configured in the Generic Pressurized Water Reactor. The present study features a plant-specific TCS with two displays side-by-side.

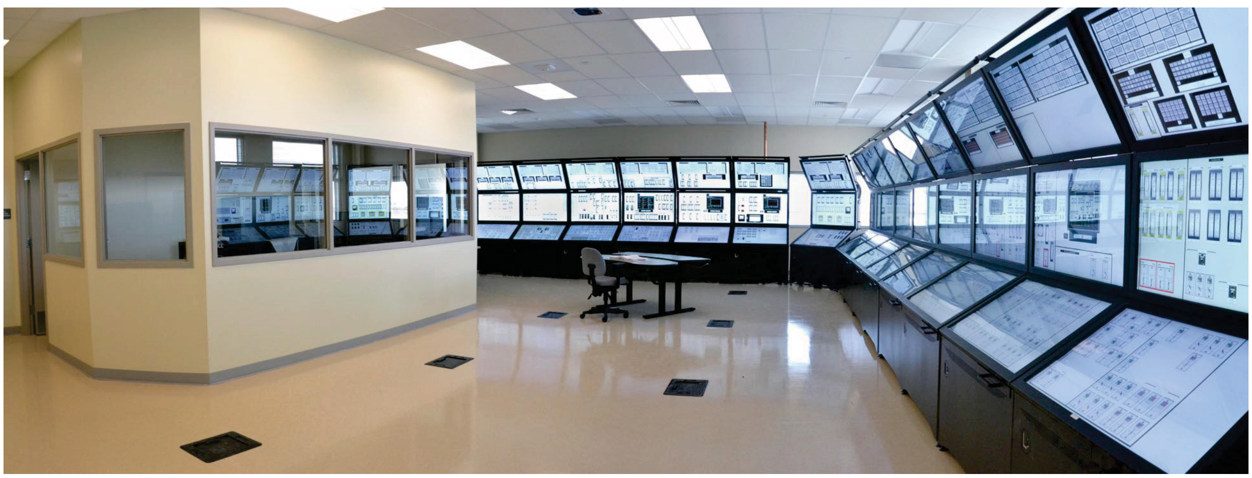

Figure 1. Human Systems Simulation Laboratory at Idaho National Laboratory.

Each plant configuration is different, and the HSSL is reconfigured to match the specific plant being studied. INL does not typically develop the glasstop renderings of the existing boards for each plant. These renderings are developed as part of the glasstop classroom trainers developed by simulator vendors for each plant. We have received glasstop versions of plant simulators from simulator vendors GSE Systems, Western Services Corporation (WSC), and L3-MAPPS, representing NPP designs by Westinghouse Corporation, General Electric, and Combustion Engineering. Additional plant simulators are typically straightforward to install in the HSSL, due to existing infrastructure developed to work with these simulator vendors.

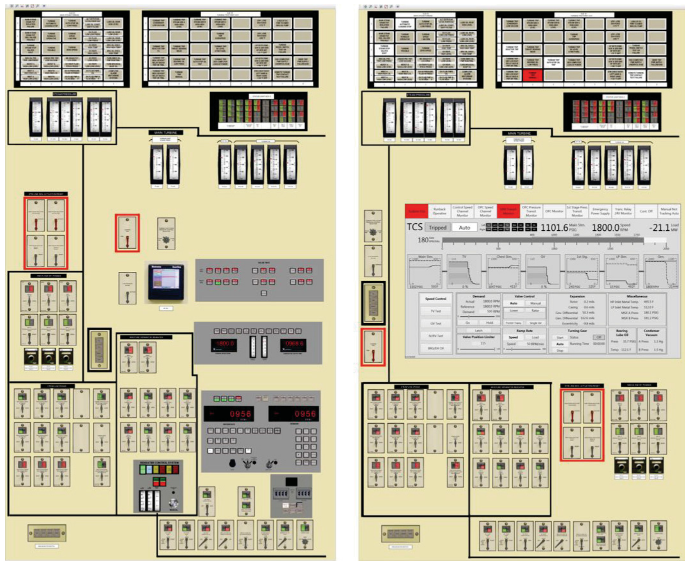

Figure 2. Example of a legacy TCS (left) and an upgraded TCS (right) in the HSSL.

When the glasstops are used for training purposes a single bay is used for the entire control room. When used in this configuration the operators can navigate from panel to panel and zoom to better observe and control I&C of interest. In the HSSL a fixed layout is assumed so the glasstops represent the layout of the control room. The objects corresponding to the I&C of the plant are moved slightly to fit the scale and dimensions of the glasstop displays. For example, rather than have a gauge span two displays vertically, including the physical bezel between the two displays, INL staff fine-tune the positioning of the objects. A gauge that spans two displays vertically, for example, would be moved upward or downward so that it only falls within a single display.

Additional adjustments are made when new virtual displays are introduced to the boards. Some I&C disappears, because it is redundant with information found on the new digital screens. Other I&C is moved to make room for the new display. In the case of OSSO-1 and -2, significant changes were made to the board associated with TCS. The basic digital TCS represents two new displays side-by-side on the boards, while the system overview represents a larger display directly above the two digital TCS displays. It is possible to subsume a large number of the existing gauges and controls through introduction of the digital HSI. This means considerable real estate on the boards is reclaimed by the upgrade, allowing additional room for new displays to support visualization. Finding room for digital HSIs on the boards can be a challenge in some modernization efforts, particularly for early vintage NPPs, which tended to have more compact control boards that are reminiscent of ship control rooms (Boring, Ulrich, and Lew, 2016). Newer plants, including the one used in OSSO-1 and -2, feature a large control room with ample room for expansion on the boards, making the prospect of digital replacement HSIs more tenable.

2.4 Prototypes

2.4.1 Development Environment

INL jointly with collaborators at University of Idaho developed the Advanced Nuclear Interface Modeling Environment (ANIME; Boring, Lew, and Ulrich, 2017). This user interface framework for Microsoft Windows Presentation Foundation allows creation of prototypes that communicate with the underlying plant simulator. The library of graphical widgets allows INL to mimic the look and functionality of a modern DCS, either closely following the conventions of a particular vendor or testing INL's own design concepts. The ANIME prototype is overlaid on the virtual control boards to represent the modernized DCS. ANIME allows rapid application development by maintaining the underlying simulator functionality. Essentially, it acts as a software skin on the control boards. For example, for TCS, the existing plant simulator contains all TCS functions. The ANIME prototype intercepts the operator input and simulator output pertaining to the TCS and presents it as an integrated digital HSI. While it maintains the underlying functionality of the modeled TCS, it also augments it. Where there are new functions proposed for the HSI, the control logic is added in ANIME. Thus, the only control logic that must be modeled is where it differs from existing logic. Keeping the existing functions "under the hood" of the prototype makes it possible to develop the prototype considerably faster than the vendor can build the first iteration of the new HSI, since the vendor must typically build all of the control logic that will form the new system.

TCS prototypes have been developed for five different plants that mimic two commercial DCS platforms. Additional systems like CVCS have also been developed into DCS prototypes. The same method has been used for early to modern digital-to-digital upgrades in the form of plant process computer screens. Initial skepticism by one TCS vendor of the value of a third-party prototype that mimicked the TCS was replaced by appreciation that issues in the design of the TCS were identified and corrected long before deployment, thereby preventing costly reworks.

2.4.2 Turbine Control System Prototype

The utility involved in OSSO-1 and -2 decided it would purchase a digital TCS from a DCS vendor. The DCS vendor prepared a detailed product specification and provided an HSI style guide to indicate the look and feel of the interface. The TCS specification covered all essential control functions such as latch, speed control, and load control; operational screens like general turbine functions, valve and trip tests, and diagnostics; required bypass functions; and system maintenance. Using the HSI style guide, previously developed TCS DCS screenshots, a standalone software demonstration of the TCS from another plant, and the specification, the ANIME development team created DCS screen mockups that matched the vendor TCS (see Figure 3). These mockups were then overlaid on the existing glasstop control boards using the Microsoft Windows topmost picture-on-picture functionality, hiding the existing analog TCS. The net result was a mimic of touchscreens placed on the plant control boards.

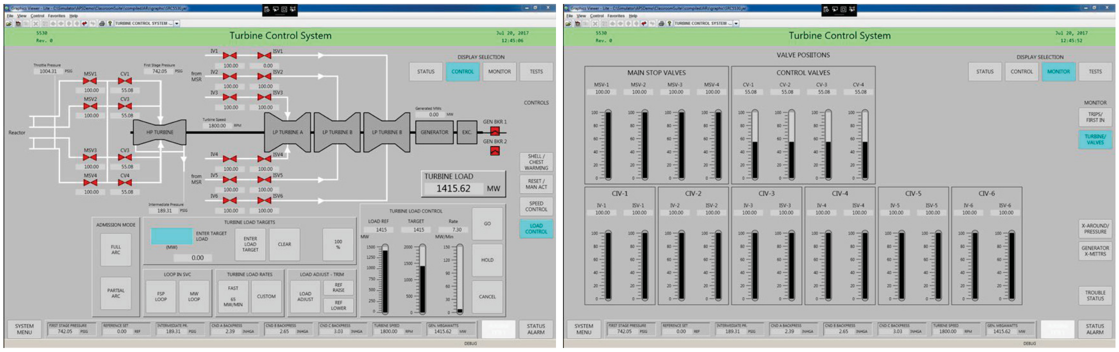

Figure 3. Example TCS prototype screens mimicking a vendor's design specification.

The prototype indicators and controls in ANIME were paired to their equivalent components in the simulator, thus keeping the existing TCS model intact on the simulator. As mentioned in the previous section, where there was new functionality added by the digital TCS, C# code and simulator scripts were incorporated in ANIME to bridge the simulator TCS control logic with the proposed digital TCS. Thus, new code was developed only where essential to model new functionality, but the underlying simulator model was not modified. The focus on difference modeling ensured rapid development times and a high level of prototype functionality without the need for plant model redevelopment. In this manner, the prototype TCS could be developed substantially faster than the actual DCS, because the prototype TCS did not need to create substantial new control logic nor meet stringent quality assurance requirements that would be required of the actual DCS.

The ANIME-based prototype was tested through a series of scenarios representative of TCS use, as described earlier. Licensed reactor operators were brought into the HSSL for operator-in-the-loop evaluation. The testing allowed operators to get hands-on experience with a close approximation of the proposed TCS and provide feedback to the design and functionality of the TCS early in the design stage.

The DCS vendor provided a standalone demonstration program of their TCS that was implemented for another plant. The rationale for developing the integrated prototype in addition to the standalone demonstration was that the prototype allows for a crew of operators to interact with all of the plant systems including the TCS. During operations, the crew must monitor and manipulate the entire plant, not just the TCS. The crew also has access to multiple TCS displays. The interaction and workload associated with the TCS may be easy to handle, but when the additional complexity of managing the whole plant is incorporated, the crew must work in concert to accomplish operational tasks. Holistic examination is needed to identify how changing the TCS affects current operational protocols and other plant systems. In addition to the need to test the TCS in the integrated context of overall plant operations, the prototype allowed development and testing of features not currently supported in the TCS, namely the overview screen. The vendor does not currently have an overview screen for TCS nor a specification for one.

These efforts can be conducted before a full specification has been defined and solutions to issues can be examined through rapid prototyping at relatively low cost. Once a fully formed TCS specification has been developed, future human factors evaluation efforts will need to integrate the actual DCS into the full plant model. This integration would require collaboration between the simulator and DCS vendor to remove the existing TCS, model new control logic, and integrate the DCS into the plant simulator build. The complexity of this task may make emulation of the DCS within the simulator environment prohibitive. In that case, the DCS becomes a stimulated system that contains its own code and communicates with the simulator. Such integration efforts are avoided during the design phase by using the ANIME prototyping tools.

2.4.3 System Overview Prototype

To test the usefulness of a system overview to support TCS operations, we developed an overview screen prototype. Note that an overview screen—as defined here—does not allow control actions. It is simply an information display of the most important parameters. In a modern control room, there are three levels of information and control that the operators must maintain: the plant level, the system level, and the control or component level. The current TCS specification provided by the vendor only encompasses control or component level screens. This study sought to introduce the system level overviews to help operators maintain situation awareness of the TCS.

While it is possible to create elaborate visualizations to trend or even infer plant parameters, the goal here in creating the overview screen for TCS was to align closely with the vendor's existing design standards. A senior reactor operator was consulted to identify the most important parameters for turbine operations, which were then extracted into design elements. Where such elements were found across the different screens of the existing TCS, they were copied as elements of the overview screen. Other elements were designed using the available graphical style elements in the DCS. Key elements included valve position, vibration, speed, load control, pressure, and temperature indicators. The initial design of the overview screen used in this study is shown in Figure 4. The system overview screen in conjunction with the two TCS control screens is shown in Figure 5.

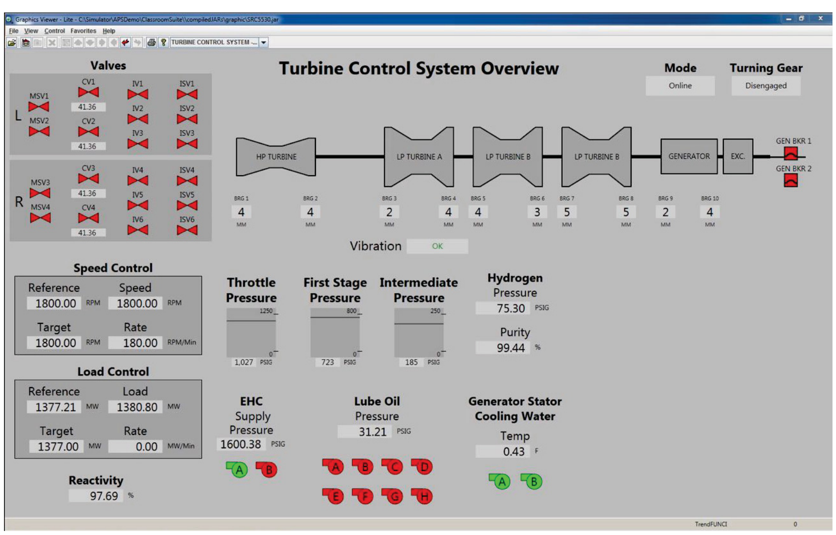

Figure 4. The turbine control system overview screen.

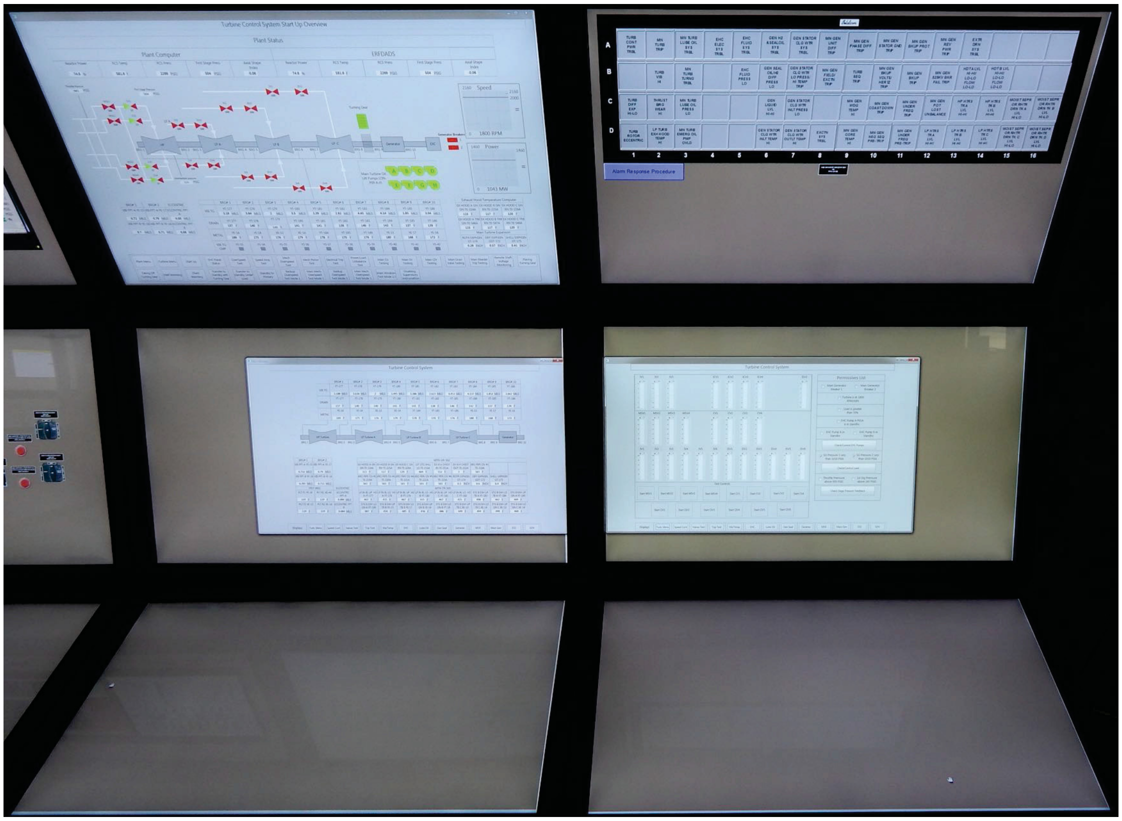

Figure 5. Example digital turbine control system control screens with overview screen in the HSSL.

Note that in the present study, there was some overlap between TCS control screens and the overview screen. In the vendor's design for the TCS control screens, many include a representation of the turbine with key indicators. This representation actually uses about a third of the screen. As such, the amount of novel information per different screen is minimal. Operators may have to navigate through multiple screens to arrive at the critical piece of information, because the amount of unique information is limited across screens. By moving the overlapping information to an overview screen, it should be possible to optimize the information presented to operators on the control screens. In the current iteration of the prototypes, the digital TCS control screens are maintained per the vendor's original specification. An optimized TCS with overview screen would likely require a redesign of the individual TCS control screens.

Some future functionality, namely a predictive fault detection system for TCS, was also tested in a few scenario variants. The results of those scenarios, which are not central to the plant's current TCS implementation, are presented in a separate report (Ulrich et al., 2017).

2.4.4 Prototype Summary

The ANIME framework developed by INL and University of Idaho researchers allows rapid application prototyping of HSIs. Because the prototype HSIs are only used for evaluation in situ on simulators within a prescribed range of scenarios, they are not subject to the same rigorous quality assurance requirements as the final deployed DCS. Additionally, because the prototypes retain existing control logic in the simulator—whereas new systems must develop entirely new control logic—it is possible to develop prototypes in a fraction of the time of the actual DCS. In this manner, it is possible rapidly to create and iterate high-fidelity prototypes with an emphasis on the HSI for evaluation studies. The advantage of prototyping is that it affords a sneak peak of system functionality early in the design phase before vendor resources have been heavily invested in constructing the DCS that will be deployed. The prototype code is considered disposable or throw-away in that it only serves to create a realistic representation of the HSI without implementing the underlying system code. The prototype HSI allows INL to provide feedback to the plant and the vendor on the design to help them finalize the system specification.

2.5 Operators

Three licensed reactor operators served as participants in the scenarios. The operators had completed two-year full-time licensing classes at the plant and worked as an on-shift crew for at least a year. The crew was divided into three roles:

- Control room supervisor (CRS)—responsible for directing control room actions, following written operating procedures, and maintaining oversight on plant activities;
- Operator at controls (OAC)—responsible primarily for controlling reactivity and associated functions at the plant; and
- Balance of plant (BOP) operator—responsible primarily for secondary side functions such as electricity generation and turbine control.

Note that the CRS does not operate the boards, while the OAC and BOP work in tandem to control the plant from the boards and may work interchangeably outside their main board responsibility areas. Absent from the crew complement in this study was a shift technical advisor (STA), who would normally assist the crew in performing tasks not requiring control actions, e.g., calculations or verification.

The study included an instructor from the plant, who administered the scenarios from the simulator instructor station. The instructor, who comes from a qualified operations background, acted to direct the scenarios at initiation and also to answer any questions requiring information outside the main control room (e.g., calls to field operators for readings from local indicators).

2.6 Performance Measures

2.6.1 Background

There were two types of data collected for each scenario:

- *Objective operator performance data*—such as task completion times and eye tracking.

- *Subjective operator performance data*—such as preference for particular systems, self-assessment of workload and situation awareness, as well as a think-aloud narrative during scenario runs.

Although various human performance measures such as eye tracking, situation awareness, and workload were collected on reactor operators using the prototypes, the purpose of gathering these measures is to help establish operator preferences and requirements relative to the new designs. This study was explicitly not to be an evaluation of operator performance but rather an evaluation of the quality and usability of the new HSIs. INL staff members previously developed the Guideline for Operational Nuclear Usability and Knowledge Elicitation (GONUKE; Boring, Ulrich, Joe, and Lew, 2015; see Table 2), which provides guidance on appropriate measures for different design phases per the U.S. Nuclear Regulatory Commission's *Human Factors Engineering Program Review Model* (NUREG-0711; O'Hara et al., 2012).

Table 2. The Guideline for Operational Nuclear Usability and Knowledge Elicitation (GONUKE), highlighting activities of the 2017 operator-in-the-loop study.

Evaluation Type	Evaluation Phase: Pre-Formative *(Planning and Analysis[1])*	Formative *(Design[1])*	Summative *(Verification and Validation[1])*	Post-Summative *(Implementation and Operation[1])*
Expert Review *(Verification)*	[1] Design Requirements Review	[2] Heuristic Evaluation	[3] System Verification	[4] Requalification against New Standards
User Study *(Validation)*	[5] Baseline Evaluation	[6] Usability Testing	[7] Integrated System Validation	[8] Operator Training
Knowledge Elicitation *(Epistemiation)*	[9] Cognitive Walkthrough (Task Analysis)	[10] Operator Feedback on Design	[11] Operator Feedback on Performance	[12] Operating Experience Reviews

[1]Corresponds to phases in NUREG-0711

Because the objective of the operator-in-the-loop study is to gather feedback from the crew to inform and finalize the design specification, much of the feedback obtained is qualitative in nature. Qualitative feedback tells the design team:

- What the operators expect the HSI to do,
- What the operators like about the HSI,
- What challenges the operators encounter while using the HSI, and
- What operators would like to see in the HSI.

This feedback is translated directly into design recommendations for the design engineering teams at the vendor and plant to consider as they finalize the specification.

We developed the concept of As Low As Reasonable Assessment (ALARA; Boring, 2016) as a way to apply discount usability to the formative stages of design evaluations. In practice, this means we do not gather performance measures we do not need to inform the design of the system. We economize the measurement in order to reduce the intrusiveness of the study to the operators and analysis time required from INL staff. Operator walkthroughs of scenarios with think-aloud protocols and verbal debriefs are typical for discount usability and are adopted under ALARA.

Additional evaluation of operators in the form of integrated system validation (ISV) occurs after the design is finalized and implemented and ensures that the design as projected actually translates into an HSI that functions safely and optimally for the operators. Using a prototype to evaluate early design concepts is considered formative evaluation, while using the actual system to validate the system is considered summative evaluation (see Figure 6). Formative performance measures tend to be more subjective and qualitative as suitable for providing design input, while summative performance measures tend to be more objective and quantitative as suitable for factory acceptance testing.

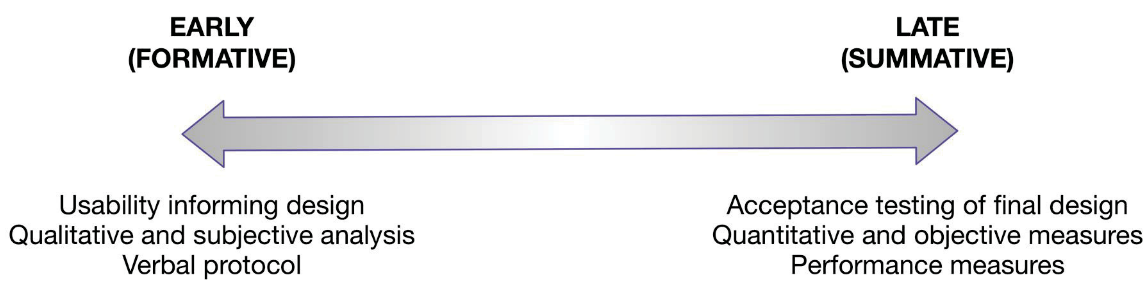

Figure 6. Evaluation continuum for design of new human-system interfaces.

As suggested by GONUKE (see Table 2), the operator-in-the-loop study features three types of data collection:

- An expert review of the TCS design by human factors experts, using the HSI style guide as a basis for the adherence to design quality conventions.
- Usability studies using the crew interacting with the full-scope simulator, prototype HSIs, and realistic plant scenarios.
- Expert feedback from the operators (called *epistemiation;* Boring, Lew, and Ulrich, 2016) on their impressions and use of the prototypes

For the usability studies, additional standard measures (see Boring, Joe, Ulrich, and Lew, 2015) were be collected for each scenario, including:

- Subjective user preference survey

- Workload measures
- Situation awareness measures
- Task timing performance measures
- Operator self-report of psychological performance shaping factors
- Audio recordings
- Video recordings
- Procedure logs
- Simulator logs
- Eye tracking
- Expert observations on performance.

These measures allowed a degree of performance quantitation to support findings from the verbal reports by the operators. While it is tempting to use these measures for objective benchmarking between different variants of the system (e.g., to compare the legacy analog and new digital HSIs), it must be noted that the quantitative measures should not be considered conclusive at this stage, because:

- Operators are encouraged to think aloud, which may slow the natural response rate to the scenarios.
- Operators have not been fully trained to use the prototype HSIs, meaning their performance reflects the discovery and learning phases of use, not the proficient phase.
- Procedures have not been rewritten to make use of the prototype HSIs.
- Because the prototype HSIs make use of existing simulator code based on the analog system, some system timing may not properly reflect the latency or responsiveness of the actual digital system.

Nonetheless, these measures do afford the opportunity to look at research questions beyond the simple operator preferences and wishes for the new digital HSIs. There are a number of complementary goals of the collaborative effort between the LWRS program and industry participants. While the operator-in-the-loop study may provide concrete answers to finalize the TCS design specification at the plant, it also provides the opportunity to generalize the results to benefit the broader nuclear industry. Thus, the results of optimizing the digital TCS should readily translate to the TCS upgrades at other utilities or, indeed, to other digital systems that are being modernized.

2.6.2 The Need for Better Performance Measures

There is considerable need to measure operational safety and human performance as new technologies are introduced into control rooms in NPPs. It is necessary to establish that operators using these technologies may perform the tasks at least as safely using new technologies such as digital HSIs as older technologies such as analog I&C. One enormous challenge is how to establish this performance given that existing plants already operate extremely safely. How can we realize safe or even better performance when we have near-perfect safety and reliability?

ISV is a common method to test operational safety. ISV serves as a type of factory acceptance test that minimum safety and usability criteria are met. ISV is important for the vendor for meeting design requirements, for the plant in establishing the system meets acceptance criteria, and for the regulator in ensuring the system meets safety standards. Simulations used in ISV—such as operator-in-the-loop studies in full-scope simulators—are usually too limited to use typical quantitative human performance measures like accuracy (i.e., human error rate) and time (i.e., performance time). Sample sizes of operators are typically too small for definitive conclusions through inferential statistics, for example. It also becomes challenging to elicit errors in operators who perform near perfectly in an existing system. It is hard to demonstrate performance improvement from near perfection. Time measures may be challenging to gather as well, since the plant limits the response time. For example, a new digital TCS to replace an analog predecessor will not necessarily show marked improvement in speed to complete a task like startup. The physics of the plant limit the ability to accelerate such processes.

Thus, several measures are used instead of accuracy and time for evaluating human performance in ISV, such as workload measures, situation awareness measures, or some subjective questionnaires. Most of the methods highly rely on expert judgment by either operators or observers. Some of the methods are not actually sensitive measures when used with experts like reactor operators. In sum, these surrogate human performance measures do not necessarily produce objective measures to demonstrate the safety of the system.

Yet, during validation studies, many raw records can be collected from simulators: audio-video records, several types of simulator logs (e.g., operator action logs, parameter logs, alarm logs, malfunction logs, etc.), and eye-tracking records. These records contain a lot of information on the plant behaviors or operator responses during simulation scenarios. For example, plant parameters such as the pressure/level of the pressurizer or steam generators or the temperature of reactor cooling system are the key indicators to show whether the plant is at a steady state. These measures may emerge from a combination of operator actions during the simulated condition. The question is how we can use the raw records of simulations for evaluating human performance. It is possible to link these raw records systematically to human performance. One goal of this study was to review the applicability of new measures of operator performance.

2.6.3 Summary of Measures

As noted, the study featured objective and subjective performance measures. Subjective survey measures included operator self-assessments of workload and situation awareness, other drivers thought to affect performance, and overall satisfaction with the system. These surveys were administered to each of the three reactor operators across the variants of the TCS, allowing a comparison between systems. In addition to the surveys, the study consisted of an extensive semi-structured debrief with the crews after every scenario run to elicit design issues and suggestions.

The standardized questionnaires are described below:

1. *Workload*—NASA Task Load Index (TLX) is a subjective multi-dimensional evaluation tool used to evaluate for perceived workload. The tool consists of six separate sub-scales:

Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. Subjects (i.e., operators) are asked to rate their workload across each sub-scale. An aggregate score of workload can be completed for general workload comparisons (Hart & Staveland, 1988). The original NASA TLX incorporates a set of 15 pairwise comparisons used to weight each sub-scale; however, the NASA TLX can also be run without these pairwise comparisons, which is sometimes defined as a 'raw TLX.' This workshop used this 'raw TLX.' The NASA TLX for this workshop used a 10-point scale (1=low, 10=high).

2. *Situation Awareness*—Situation Awareness Rating Technique (SART) is a subjective multi-dimensional evaluation tool used to evaluate perceived situation awareness (Taylor, 1990). For this study, a total of nine of the ten questions were used (i.e., omitting the question arousal level), each using a 10-point scale (1=low, 10=high). Each question was evaluated separately.
3. *Other Performance Drivers*—A final questionnaire assessed operator self-assessment on several performance shaping factors (PSFs) derived from the human reliability analysis (HRA) literature. A PSF is anything that influences human reliability. PSFs are included in most HRA methods, ranging from cognitive factors to task and situational factors to aspects of the environment. NUREG-1792, *Good Practices for Implementing Human Reliability Analysis* (Kolaczkowski et al., 2005), provides a standard list of PSFs. This list is easily mapped to the lists of specific PSFs found in HRA methods, although different HRA methods may emphasize other PSFs or classify the factors into different taxonomies. Recent work on HRA for validation (e.g., Boring, 2015; Boring, Blackman, and Rasmussen, 2017; MacLeod, 2015) suggests that human factors studies may rely too heavily on measures like workload and situation awareness while ignoring other indicators of performance. The field of HRA has catalogued dozens of PSFs that are known to cause detriments in operator performance.
4. *Overall Impressions*—Questionnaires asking about overall impressions of the HSIs. These questionnaires contained questions using 5-point Likert scales, as well as open-ended questions for qualitative analysis. Likert scales used common adverb intensifiers as shown below in Table 3.

Table 3. Likert scale used for impression survey.

Strongly Agree	Somewhat Agree	Neither Agree nor Disagree	Somewhat Disagree	Strongly Disagree
1	2	3	4	5

The questionnaires do not directly drive design recommendations but are helpful to determine operator preferences as well as workload and situation awareness. It should be noted that the operators were using the new HSIs in a novice user mode without calibrated procedures to guide operation of the digital controls. As such, these results represent their first exposure to the new digital HSIs, not their impressions and performance after extensive training and experience. The fact that operators in some cases preferred the new HSIs at first use is encouraging, but it is important that evaluation should be ongoing. As operators have more experience with digital HSIs, it's important to re-evaluate their preferences, impressions, and performance. Such latter data points are the true litmus test of the usability of the digital systems.

Open-ended questions were presented verbally to the crew to elicit their impressions of the HSI. These questions did not have specific quantitative scales. Instead, they were designed to trigger responses by the operators on what they liked or did not like about the HSIs. These open-ended questions became the basis for design recommendations on the TCS.

Eye tracking was used to assess attention while the operators completed various scenarios. Each operator wore a glasses-based mobile eye tracking platform capable of capturing fixation and saccade data at 60 Hz for both eyes. This eye tracking unit also video recorded the visual scene of each operator to support data analysis. To process the eye tracking data, the included SMI software suite was used to map the fixations recorded within the three-dimensional visual scene of the eye tracking glasses onto the flattened reference images of the control boards. This time consuming process was necessary to perform statistical analysis of the raw eye tracking data in order to generate eye tracking metrics (Kovesdi et al., 2017). Two primary metrics were used, which both pertain to fixations. The first was the relative number of fixations within each area of interest. Second, heat maps and scan patterns were created to provide visual representations of attention patterns and further explain the proportions of attention within each area of interest. The primary hypothesis evaluated with eye tracking entailed comparing the patterns of attention between the analog, digital, and digital with overview screen configurations. Operators were expected to have a more concentrated pattern of attention for the digital TCS screens than the analog display configurations.

Finally, a series of plant parameters were logged from the simulator. These data points were sampled every second. The parameters corresponded to parameters found in the SPDS (corresponding to the main indicators of plant functioning) and the TCS overview screen (corresponding to the main indicators of TCS functioning). Some of these parameters are presented in a preliminary analysis in this report. Recent work at INL in the area of human reliability analysis has linked objective plant parameters to human performance metrics, suggesting it may be possible to quantify the influence of the particular plant contexts on operator reliability (Rasmussen et al., 2017). INL will make use of the parameter logs from this study to help calibrate the link between plant states and operators in future research on human reliability analysis.

Although audio and video logs of each scenario were recorded, the logs were not analyzed for this report. The logs were simply used to reference segments of interest during the scenario or capture helpful comments by the operators. Additional post hoc analysis of crew interactions are possible in the future, e.g., to establish levels of mutual awareness among operators when using different technologies (Savchenko et al., 2017).

2.7 Study Procedure

Each participant in the study completed informed consent forms prior to the study. Per the informed consent, participant identities and performance are anonymized in all reports of the study. At the beginning of each scenario, the instructor provided a brief description of the initial plant state and required task (e.g., “Follow the turbine startup procedure until further instructed”). A visual and auditory start cue was revealed, at which point the simulator was placed into run mode and note-taking software was started. The visual and auditory cue allows synchronization of plant logs, observer notes, eye tracking, and the audio and video recordings. The operators proceeded through the scenario uninterrupted,[3] carrying out standard threeway crew communications. The end of each scenario (e.g., the point at which the turbine was successfully synchronized to the grid) was specified in the detailed scenario instruction scripts. Upon reaching the specified stop point, the instructor notified the crew that the scenario objectives had been met and that the simulator was being put into freeze mode. Following this, the crews were given the questionnaires described in Section 2.6.3. Upon completion of the questionnaires, the study facilitator asked debrief questions designed to elicit operator impressions of the HSI. Observers took notes to capture the points made by operators during the debrief. The debrief was conducted in a group setting, but following a group discussion, each operator was asked for their individual response in a roundtable fashion. This process ensured that each operator provided feedback to each question asked during the debrief.

[3] Due to a simulator fault in one scenario and a bug in one of the prototypes in another scenario, two scenarios had to be interrupted or restarted. These scenarios were excluded from the analyses.

(This page intentionally left blank)

3. STUDY FINDINGS

3.1 Objective Performance Data

3.1.1 Task Completion Time

Figure 7 illustrates the time required to complete the basic activities of turbine startup. The elapsed time is representative of the activities taken to a common procedure step, making it possible directly to compare the conventional analog boards, digital TCS, and digital TCS with overview screen. The startup scenario required 42 minutes and 49 seconds on the analog boards, 38 minutes and 24 seconds on the digital TCS, and 34 minutes and 44 seconds on the digital TCS with overview screen. By design, the overview screen included many of the key indicators required during startup, which would be the most monitoring intensive evolution for the turbine. As such, it was hypothesized that the overview would facilitate and ultimately accelerate the process of monitoring key turbine parameters.

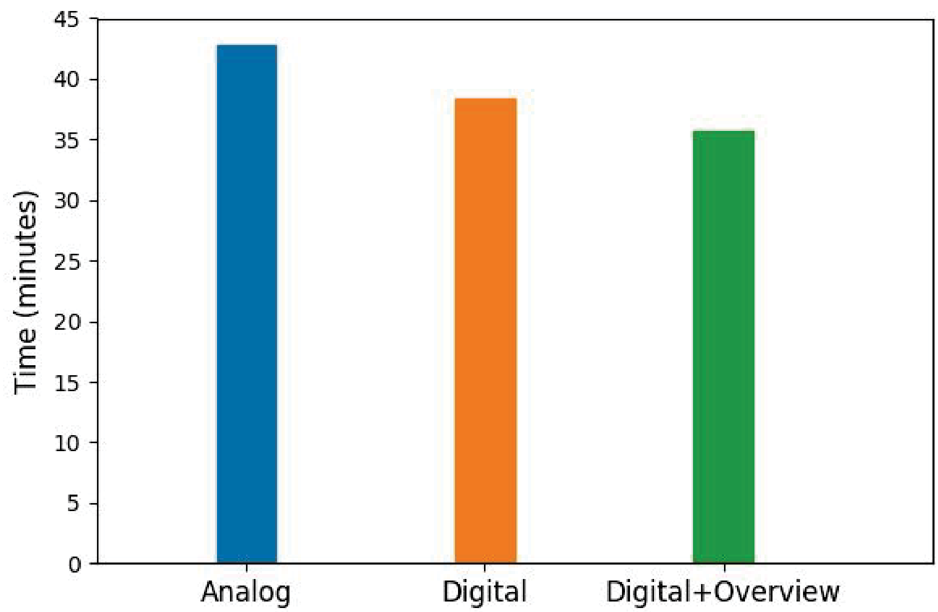

Figure 7. Comparison of startup task completion times for analog vs. digital TCS vs. digital TCS with overview display.

Because it was not possible to counter-balance the order of the startup scenarios, it is possible that there is a practice effect—that operators became quicker with each successive scenario run. We acknowledge this as a possible confound, and it is not possible to say definitively that the digital TCS and digital TCS with overview resulted in faster startup operations. However, it must be acknowledged that the crews performing the startup scenario were trained and rehearsed to high proficiency on startup at the plant. It is therefore not expected that the operators would show marked improvement or acceleration of their performance just through practice of the scenario. In fact, their proficiency at using the existing analog system might arguably result in negative transfer—in which their familiarity with one system could lead to difficulty completing the task in a new way, which could actually slow performance. The topics of practice effects and negative transfer between the HSIs remain questions for further research.

3.1.2 Plant Parameters

Operators completed a load reduction scenario using the traditional analog control boards and the new digital TCS.[4] The load reduction scenario initiated with the plant at 100% power generating around 1400 MW. The operators were instructed to reduce load by 30%. The simulator logs were examined to compare key parameters between the traditional analog and new digital ovation system (see Figure 8). There is a noticeable difference in the patterns for each parameter, in which the ovation system was able to provide a smoother transient as the control valves were closed during the lead reduction procedure. Ideally, the load reduction would follow a linear trend in which the parameters smoothly change to their corresponding value following the load reduction. This smooth transition is referred to as a "bumpless" transfer.

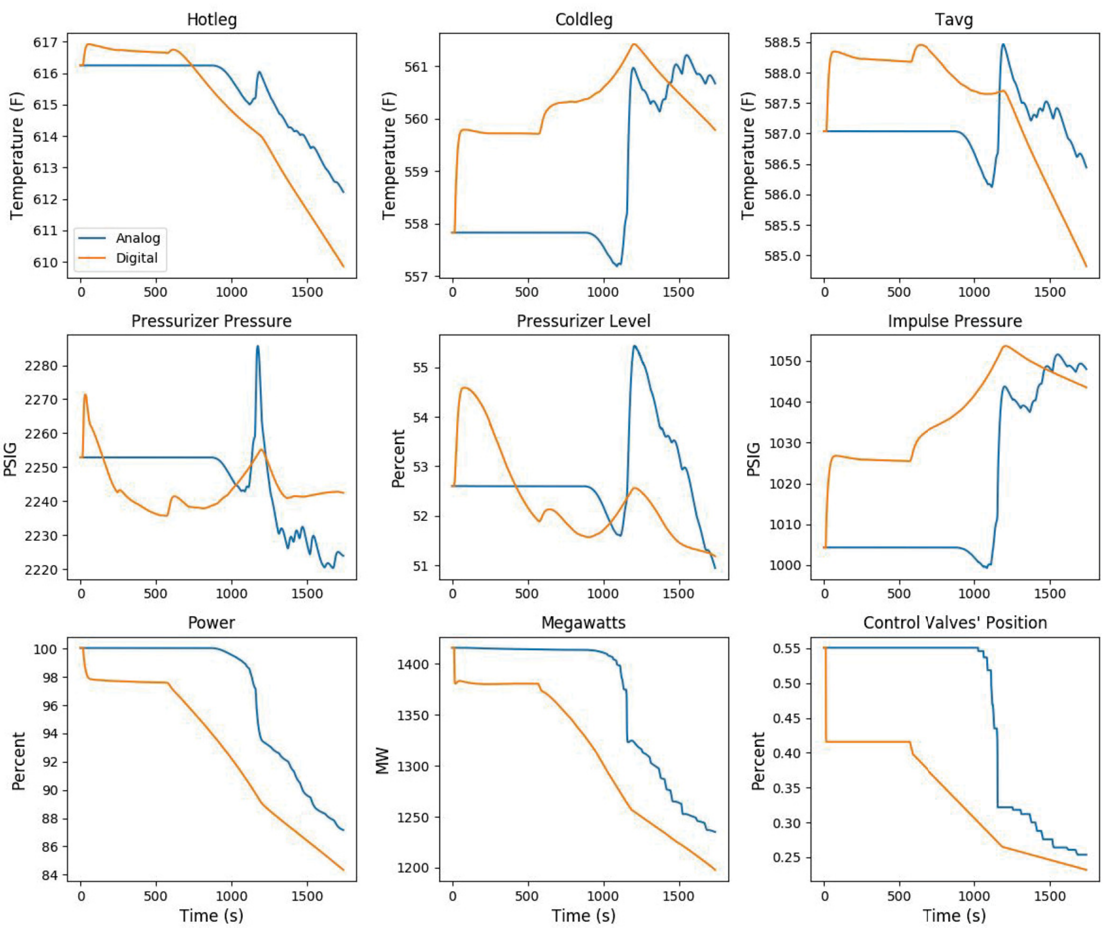

Figure 8. Key process parameters for the analog and digital interfaces recorded during the load reduction scenario.

[4] Due to a system fault, no load reduction scenario was completed for the digital TCS with overview screen.

To analyze the two interfaces within the context of these fluctuating key parameters and quantify the error due to system perturbations, the residuals of trend lines for each parameter were calculated. The root mean square (RMS) was then calculated for the residuals to yield a measure of error associated with each interface's control ability. With the exception of the Hotleg and Average Temperature (Tavg) parameters, the digital HSI demonstrated less noise, which indicates superior performance (see Table 4).

Table 4. Root mean square of the residuals for each parameter controlled using the analog and digital human-system interfaces.

Variable	Analog	Digital
Hotleg	0.61	0.74
Coldleg	0.86	0.46
Tavg	0.40	0.55
Pressurizer Pressure	8.70	6.69
Pressurizer Level	0.88	0.59
Impulse Pressure	11.15	4.70
Power	2.32	1.13
Megawatts	33.43	16.96
Control Valves' Position	0.06	0.02

3.1.3 Eye Tracking Results

Heat maps were generated for the three HSIs for the startup scenarios as shown in Figure 9. The heat maps were generated from the BOP operator, who was responsible for control of the turbine. The heat map for the analog control board reveals fixations primarily on key indicators and controls, which presumably are associated with very specific monitoring and control functions required in the turbine startup procedure. For the HSI with the digital TCS, much of the dwell time is centered on the two side-by-side TCS screens, with some additional focus areas on supporting controls.

For the digital TCS with the overview screen, the attention is divided more evenly between the three screens, although some areas of the main digital TCS are used less, presumably since they are somewhat redundant to the overview screen. The heat map for the digital TCS with the overview screen also features a scan pattern that more closely mimics the analog boards in that the operator was referencing very specific indicators and controls. Figure 10 catalogs different areas of interest (AOIs) for the digital TCS and digital TCS with overview. These are summarized for easier comparison in Figure 11. The addition of the overview resulted in less time spent looking at the alarm boards, more time looking at main steam indicators on the boards, and less time directly looking at the digital TCS screens.

The reason for the broadening of the scan patterns with the addition of the overview screen may be due to the overview screen drawing attention to specific information and the operator then verifying that information. This is a common procedure with extant digital HSIs at the plant, whereby a verification process is employed to confirm values are the same between analog and

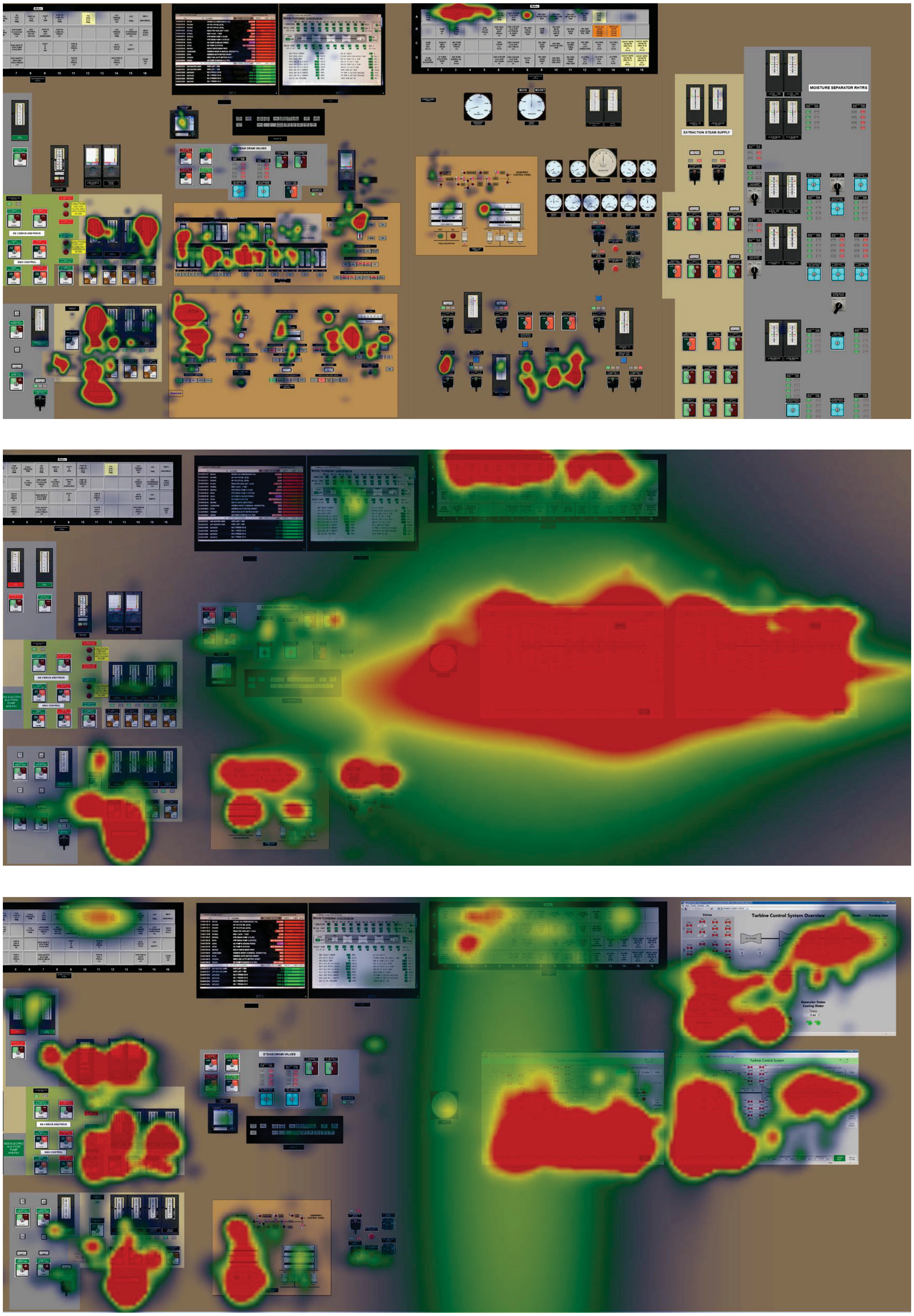

Figure 9. Heat map for analog (top) vs. digital TCS (middle) vs. digital TCS with overview display (bottom).

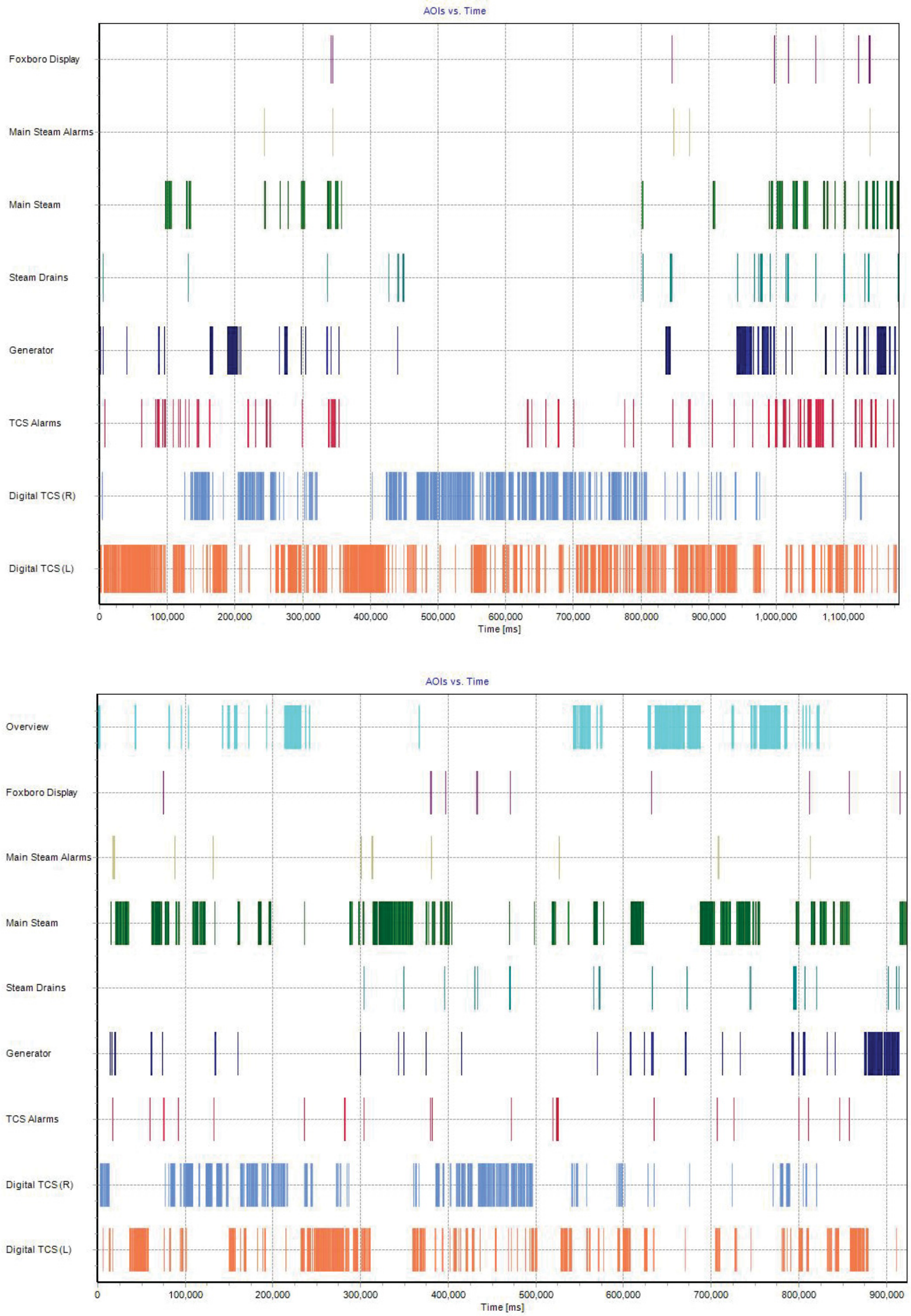

Figure 10. Areas of interest (AOI) over time for digital TCS (top) vs. digital TCS with overview display (bottom).

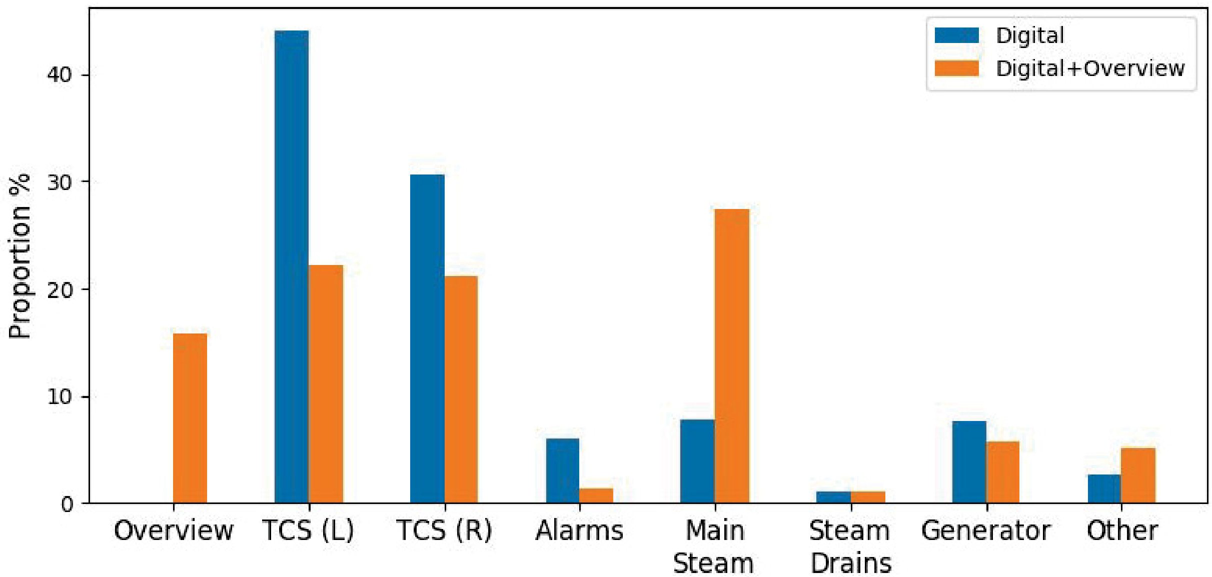

Figure 11. Proportions of attention across different systems for digital TCS vs. digital TCS with overview display.

digital indicators. Some digital values are not qualified values, and the operating license and procedures require use of qualified values in the form of the analog indicators.

Tunnel vision or the so-called Keyhole Effect (Woods, 1995) is a documented concern with digital HSIs, whereby operators may fixate on specific pieces of information on the display without maintaining broad situation awareness. Because digital HSIs may nest information in the form of multiple windows or screens, the Keyhole Effect potentially increases the likelihood of missing important parameter information that is not visible. A principle advantage of analog boards is that all information is typically available in parallel at a glance. Digital technology affords the opportunity to aggregate information into more readily understood arrangements, but limited screen real estate often constrains the ability to display as much information as would be found in the comparable analog system. Scanning the boards is a necessity of analog control boards, but it is also a desirable characteristic to ensure operators maintain cross-system situation awareness. A goal of overview screens is to help maintain situation awareness in operators by providing key information at a glance. Overview screens attempt to capture the best of analog parallel and digital optimized presentation of information. The broader scan patterns of the operator while using the overview screen may suggest that the operator is, in fact, using the overview screen to maintain awareness of the system beyond the information provided on the main TCS screens.

3.2 Operator Self-Report Measures

3.2.1 Workload: NASA TLX

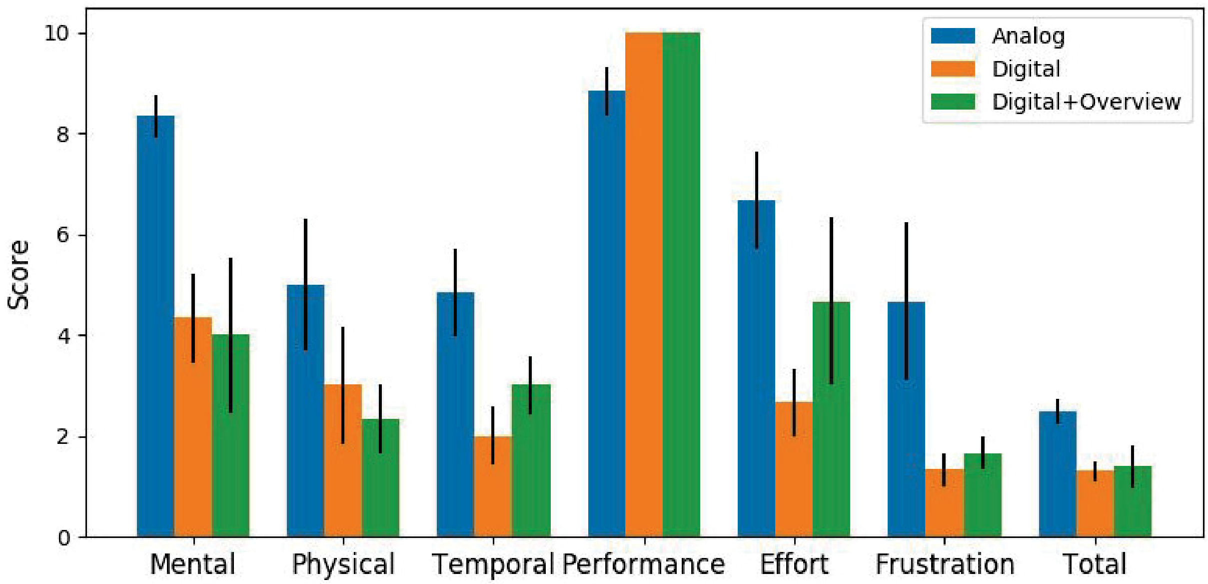

Figure 12. NASA TLX scores for analog, digital, and digital with overview configurations.

Figure 12 illustrates raw NASA TLX workload self-report scores for the three different variants of the TCS across the six operators. Standard error bars are provided to show the variability between operators. The absence of error bars suggests all operators scored the item identically. The six workload dimensions are aggregated into a total category on the rightmost area of the figure. Five of the six dimensions (i.e., Mental, Physical, Temporal, Effort, and Frustration) are calibrated such that a higher score corresponds to higher mental workload. Thus, it is generally desirable to have lower scores for these dimensions, corresponding to lower workload. One dimension (i.e., Performance) is reverse scored such that a higher score corresponds to higher performance. For this dimension, it is desirable to have a higher score, indicating higher self-assessment of personal performance. To calculate the total score, the polarity of the performance measure was reversed to correspond to the other scale dimensions and allow aggregation of all six dimensions. The total score is the average of the six dimensions. The total score is pulled down by the extreme scores and reverse polarity of the Performance dimension.

As can be seen, in no case is there are no significant differences between the standalone digital and digital with overview HSI conditions. On individual dimensions, there is evidence to suggest some difference between the analog and digital (i.e., both digital TCS and digital TCS with overview) HSIs for Mental Workload (i.e., amount of mental effort required), $F(2,4) = 4.423$, $p = 0.097$. There is a marginally significant overall effect, $F(2,4) = 5.736$, $p = 0.067$, suggesting there was higher workload for using the analog HSI vs. the digital replacements. While p is greater than 0.05, it should be noted that the small sample size severely limits statistical power and increases the likelihood of Type II error (i.e., not finding an effect when an effect truly exists).

All other dimensions were nonsignificant. While there is a marginally significant difference, it must be noted that the workload measures are generally low in all conditions, and this finding should be seen as a decrease on generally low workload for the analog TCS. It is also worth noting that the study featured a very small sample size of three operators, providing very little statistical power to establish significant differences.

3.2.2 Situation Awareness: SART

SART is a multidimensional post-trial subjective rating technique used to quantify an individual's situation awareness (Taylor, 1990). The traditional SART consists of the 10 dimensions, which Taylor refers to as generic situational awareness constructs (see Table 5). These ten dimensions were identified through a principle component analysis based on the ratings of situations presented to pilots. Initially, a group of pilots were interviewed to generate a list of 29 scenarios in which half of the scenarios were characterized as low awareness flight scenarios and the other half were high awareness flight scenarios. For example, a low awareness flight scenario was "flying in formation in an unfamiliar aircraft working at the limit of your capacity" and a high awareness flight scenario was "approaching to land in good weather at a familiar airfield, in a familiar aircraft fitted with good displays" (Taylor, 1990, pg. 3-3).

Table 5. The ten dimensions of the SART placed into the three broad categories of demand on attentional resources, supply of attentional resources, and understanding of the situation (adapted from Taylor, 1990).

Generic Construct	Dimension	Description
Demands on Attentional Resources		
Instability	Unstable vs. stable situation	Likeliness of situation to change suddenly
Complexity	Simple vs. complex situation	Degree of complication (number of closely connected parts) of situation
Variability	Few vs. a lot of things to attend to	Number of variables which require one's attention
Supply of Attentional Resources		
Arousal	Low level vs. high level of arousal	Degree to which one is ready for activity (sensory excitability)
Concentration	Low level vs. high level of concentration	Degree to which one's thoughts are brought to bear on the situation
Focusing	Focused vs. divided attention	Degree of distribution of focusing of one's perceptive abilities
Spare capacity	No vs. a lot of spare capacity	Amount of mental ability available to apply to new variables
Understanding of Situation		
Information quality	Poor vs. good quality of information	Degree of goodness or value of knowledge communicated
Information quality	Poor vs. good quality of information	Amount of knowledge received and understood
Familiarity	Unfamiliar vs. familiar situation	Degree of acquaintance with situation experience

A repertory grid triadic method of presentation was used to elicit constructs from a separate group of pilots presented with the 29 scenarios. The scenarios were presented in triads and each pilot was tasked with identifying two scenarios that contained something important to situation awareness that the other scenario did not contain. The pilots were then asked to explain the discriminating characteristics of the scenario and provide a situation awareness related construct to represent the characteristics. Through this process a total of 44 constructs were identified. These 44 constructs underwent a principle component analysis that was used to select the best 10 constructs based on elicitation frequency, strength of component loading, and inter-correlation clustering. The ten dimensions can be summarized by three general constructs of demand on attentional resources, supply of attentional resources, and understanding of the situation.
The SART form itself uses a polar rating scale in which an individual rates their experience after performing a scenario along each dimension. A rater tabulates the scores and calculates an overall situation awareness (SA) score with the following equation:

$$SA = \text{Understanding} - (\text{Attention Demand} - \text{Attention Supply})$$

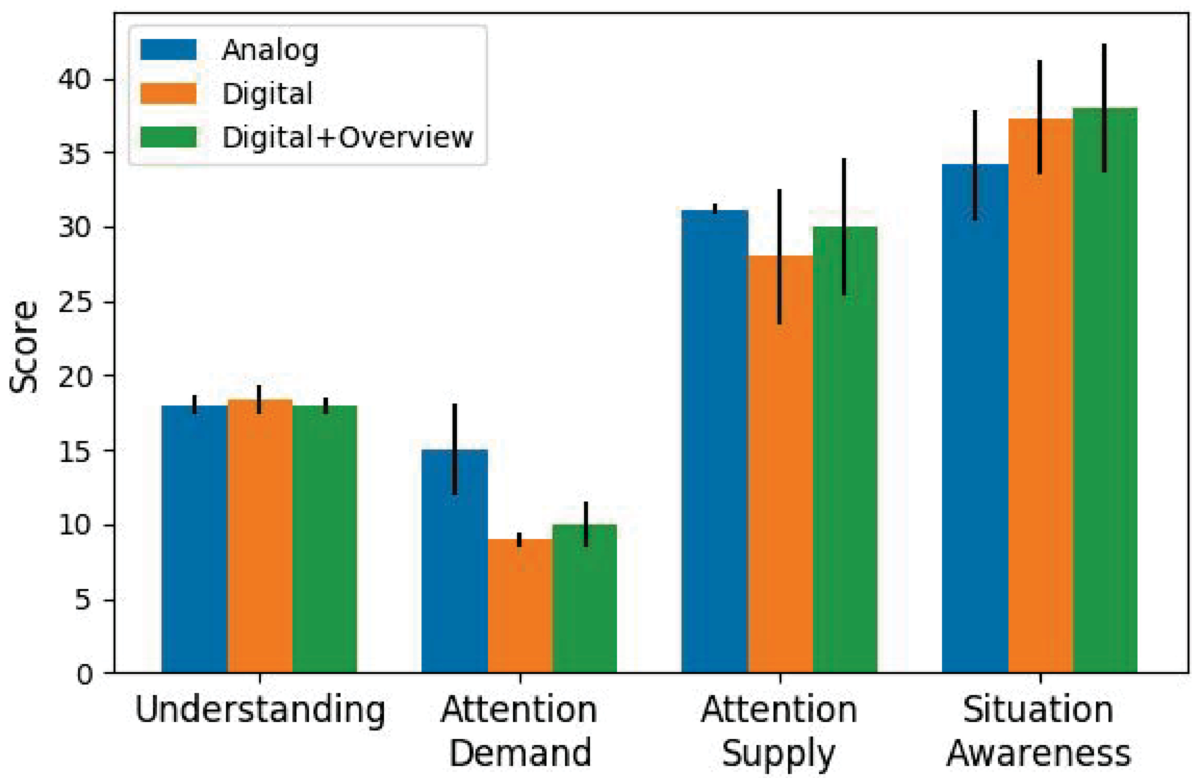

Figure 13. SART scores for analog, digital, and digital with overview configurations.

As shown in Figure 13, there is no marked or significant difference for situation awareness across the three conditions, either for the individual dimensions or overall. Overall, situation awareness is slightly lower for the analog than for the digital HSIs, but not in any meaningful way. While some variability in situation awareness might be expected when changing the manner of presenting information, it remains an important finding that the digital HSIs did not noticeably decrease situation awareness compared to the existing analog TCS.

3.2.3 Other Drivers

Workload and situation awareness continue to be key measures in human factors, but they are not the only measures of operator performance (Boring, 2015). Figure 14 considers 12 additional drivers on operator performance. This list of performance drivers is presented as a preliminary glimpse of other ways to measure performance, but the scale is still under development and has not yet been fully validated. The factors are derived from common measures that are known to influence human reliability and that are commonly considered in risk analyses (Kolaczkowski et al., 2005). The scale used for each dimension is a five-point subjective self-assessment Likert scale, ranging from -2 when a factor is deemed to have a highly negative influence on performance, to 0 for no effect on performance, to +2 when a factor is deemed to have a highly positive influence on performance. The factors are not truly orthogonal, and it is expected to see some correlations among the measures. As in previous figures, error bars are displayed where there is variability between operator responses. The results represent the average across the three operators. There were no statistically significant differences overall or for individual drivers in the questionnaire.

Almost all factors were considered positive drivers on performance, although some were more positive than others. For example, the crew self-reported Team Dynamics, Experience, Training, and Scenario Complexity as highly positive drivers on performance, regardless of the HSI. The first three factors in that list encompass personal factors and reflect confidence in the ability of the operators to perform the tasks individually and as a team. Scenario Complexity is a task characteristic. The high score on this factor suggests the crew felt confident that they had good mastery over the task and that the scenario was not overly complex.

In contrast, several factors were rated as neutral or low positive drivers on performance, including Procedural Guidance, Stress, and Time Pressure. These factors may be considered areas where there is minimal effect on the outcome. For example, Stress and Time Pressure—two situational characteristics—can be interpreted not to have played a strong role on performance. The procedures, which are certainly critical to the execution of the task, were not deemed important drivers on performance by the operators. This does not rule out that procedures could play a role in some scenarios, but the procedures were adequate for the task. This is an important finding because the procedures were not adapted from the analog TCS to the digital TCS variants. Operators did not perceive this as a stumbling block.

There are a few cases where there are differences across the HSIs. For example, for Execution Complexity and Adequacy of Time, the analog TCS was less positive as a driver than the two digital TCS variants. This finding suggests the operators felt the task easier to execute and with less time pressure for the digital HSIs than the analog HSIs. For the Indicators and Conditions driver, the analog TCS averaged close to neutral, while the digital TCS with overview actually scored as a slightly negative driver. The digital TCS without overview was considered a positive driver. The reason for this difference may be a result that the indicator for the turning gear was unreliable on the overview screen. The negative score may reflect a lack of trust in the indicators or a lack of familiarity with the overview concept as supplemental information.

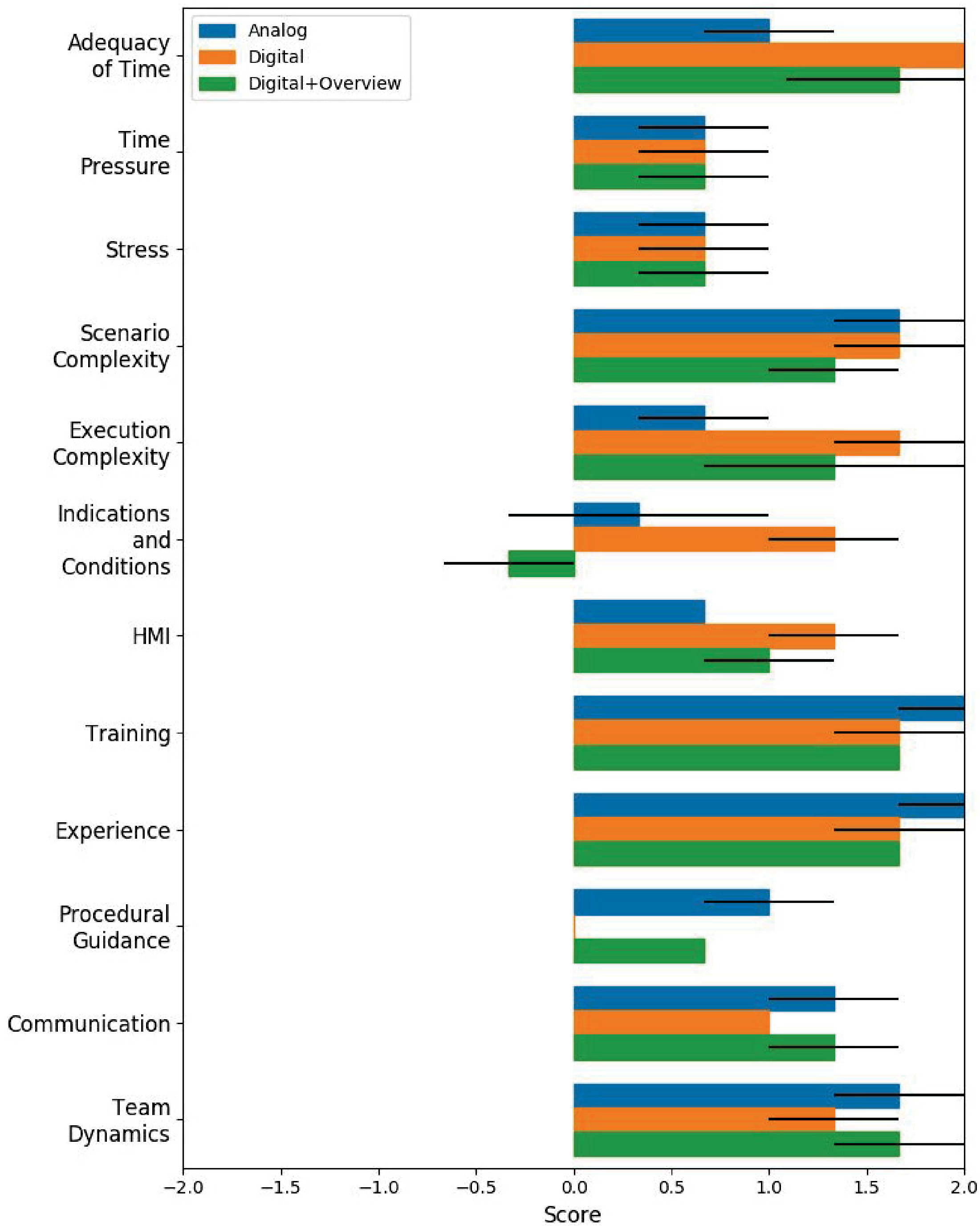

Figure 14. Self-reported performance drivers for analog, digital, and digital with overview configurations.

3.2.4 Overall Impressions

Operators were asked their agreement on three questions related to their overall impressions of the three HSI variants:

- The HSI works well,
- The HSI operates without any problems, and
- The operators like the HSI.

As shown in Figure 15, operators generally preferred the existing analog TCS over the new variants (as denoted by higher scores), although they trusted all three variants of the TCS to operate without problems. Since no distaste was expressed by the operators for the digital HSIs, it might be assumed that familiarity with the analog TCS led to generally higher scores.

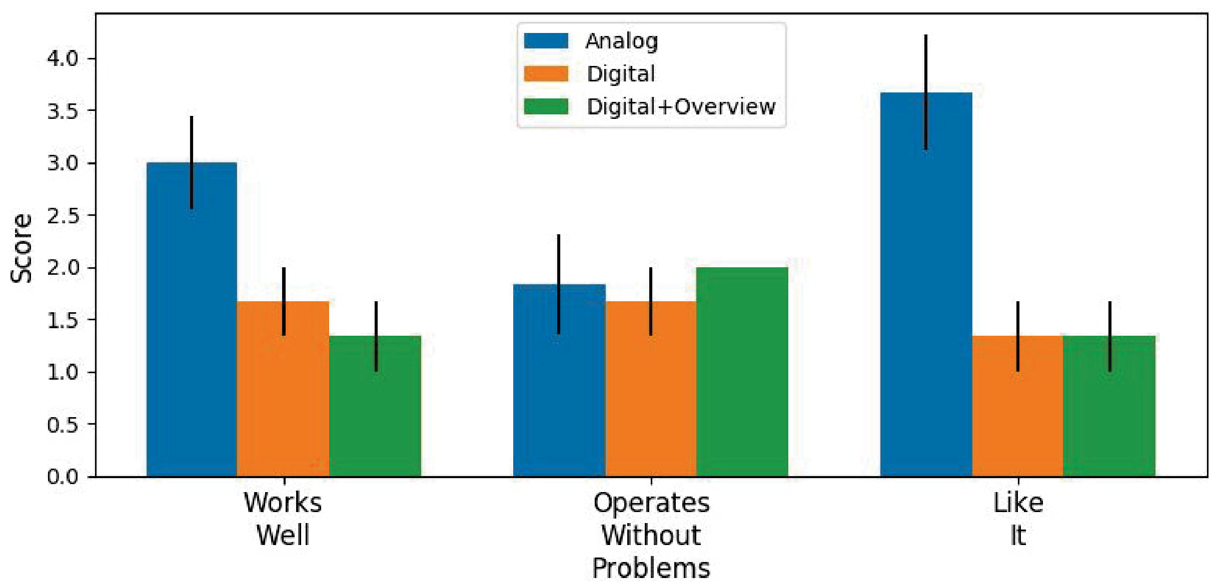

Figure 15. Overall impressions for analog, digital, and digital with overview configurations.

3.3 Findings from Operator Debriefs

Most of the findings from the operator debriefs refer to specific screens linked to the particular plant or DCS vendor. These comments are not provided here, since they are plant-specific. General trends for such comments include:

- Mismatches in terminology between the existing analog TCS and the new digital TCS. For example, the new TCS uses the term *reference* while the existing analog TCS uses the term *demand.*
- Values for predefined selections like hold points were not calibrated to the plant.
- Some fine adjustments such as those done with dial controls may not readily translate to digital setpoints, and the TCS may need more functionality to more readily fine-tune values during operation.
- Some units of measure are not calibrated to the plant.
- Existing procedure steps do not clearly map to the new digital TCS.

In short, the DCS, which is taken from a generic solution by the vendor, needs to be localized to the plant in terms of nomenclature, setpoints, adjustments, and units. Unnecessary or unintentional changes to the existing TCS are potential error traps that interfere with existing expertise in operating the plant. Additionally, procedures need to be updated to reflect the new system.

Operators expressed some frustrations with the current analog TCS, including:

- The need to enlist additional reactor operators to perform startup, since it requires two operators just to control the TCS, potentially leaving other plant functions neglected.
- The need for frequent replacement of some potentiometers to maintain precision adjustment.
- The need to adjust multiple controls in parallel to change the ramp rate or load.
- The lack of trending indicators for key parameters.
- The need to hold buttons to actuate certain activities like changing load with the load motor or valve testing.
- Startup checks are time consuming and often involve waiting for field operator response.

Operators liked several aspects of the digital HSI, including:

- Intuitive interface.
- The right information and controls for specific TCS tasks (i.e., screen tailored to specific functions).
- Automatic picking up of load adjustment after synchronization to grid.
- Stable, programmable load changes compared to the discrete adjustments (sawtooth functions) when adjustments are performed manually on the analog TCS by adjusting one parameter and then another to bring about the overall effect.

In addition to the generic enhancements described at the beginning of this section, operators suggested several potential enhancements to the digital HSI, including:

- Multiple display configuration will allow for turbine mimics to be available on overview. The screen real estate on the control screens could be modified to include task relevant indications.

- Addition of quick-button actions for time-critical activities.
- Addition of confirmation dialogs for actions that potentially change the turbine function.
- Additional trending capabilities for key parameters.
- Additional automation such as startup checks.
- Addition of embedded field indicators to eliminate need for field operator checking.
- Addition of load rate trim buttons to fine tune load rate while in Go.

Operators expressed similar comments for the digital TCS with the overview screen. They liked several features of the overview screens, including:

- The ability to see the most important parameters at a glance without the need to navigate between control screens.
- The ability to see the overview from across the room, whereas the digital TCS displays require proximity to see them properly. The increased visibility is due to the larger display and the location higher on the boards, so the display is not occluded when other operators stand in front of it. This advantage is particularly helpful to the CRS, who can more easily monitor operator actions on the TCS with an overview screen.
- Easier coordination of information between multiple operators and the CRS.
- In one variant of the overview screens, the highlighting of the affected component during alarm states. This feature showed as a red indicator next to an alarmed indication, which drew attention to the indication and reminded operators what to monitor during fault situations.

The operators expressed concern about a faulty indicator (i.e., turning gear status) in the overview screen. Additionally, although they generally agreed on the types of indicators provided in the overview screen, some specific parameters were not optimized for the plant. For example, there are multiple vibration indicators available for the turbine. The specific vibration indicators selected did not represent the indicators the operators would have chosen. So, it was the right concept but the wrong implementation.

Operators noted some potential enhancements for the overview screen, including:

- Mapping of the screen to the procedure to allow quick cross-checks. Not all parameters mentioned in the procedures are included in the overview screen.
- Elimination of redundancy between the control and overview screens. As previously noted, the vendor TCS screens were not modified for the present study to complement the information found in the overview screens.
- Additional trending options, including the capability for custom trends on the overview screen.
- Ability to tell plant mode of operation and customize information content accordingly.
- Quick checks of secondary parameters during startup.

Overall, the operators responded positively to the digital HSI variants, and saw increased efficiency with the addition of the standalone digital TCS and increased ease of monitoring with the addition of the TCS overview screen. The digital TCS would benefit from localization to the plant and from reduction of redundant information that would be found in the overview screens.

4. DISCUSSION

4.1 Key Findings on Digital Turbine Control System

While the findings may fall short of recommending a particular variant of the TCS, the study revealed that operators were able to complete the turbine scenarios successfully with all three systems. There are some reported deficiencies with the analog TCS, notably the need to adjust multiple parameters concurrently and the need to have multiple operators dedicated to the turbine for tasks like startup. The operators successfully performed the turbine tasks on the digital TCS variants without training and without tailored procedures, even affording a slight speed advantage for the new systems. The operators also appeared to benefit from the addition of the overview screen, demonstrating improved indicator tracking and broader scanning of the boards compared to the standalone digital TCS. Despite the successful performance across all three HSIs, the operators did not generally report improved workload or situation awareness, nor did they prefer the new TCS over their current TCS. This evidence suggests that initial results for the new TCS and overview are promising, but it will take additional training and design iterations for the operators to feel completely comfortable with the new TCS. Part of user-centered design processes like GONUKE is iteration, and the results from this formative evaluation paint a promising trajectory as the digital TCS and overview screen continue to be refined and implemented.

4.2 The Performance-Preference Paradox

The findings are not conclusive in terms of the degree to which the overview display enhanced the operator experience with the TCS. While there is clear indication that time to increase the task improved with the digital HSIs and that the digital variants caused smoother turbine transitions, operators generally preferred the existing analog TCS. This phenomenon is well known in the field of usability engineering, where a cardinal rule involves watching what users do and not what they say they do or want. It is the job of the human factors researcher to be the advocate for the user, which sometimes means balancing user desires or preferences with realistic user needs (Kaufman, Rojas, and Mayer, 1993). Here we refer to the present findings as the Performance-Preference Paradox—where operator preference for a system does not match their performance. We made a similar discovery when investigating input devices for a DCS among a different group of operators (Ulrich, Boring, and Lew, 2015). The operators preferred the touchscreen mode of input, but they actually performed best using a mouse. In fact, the touchscreen was prone to inadvertent input errors—a fact that would preclude that device from deployment without additional operational safety checks.

In the case of the present study, it is likely that the familiarity with the existing analog TCS translated into higher user satisfaction with that system compared to the novelty of the two new digital TCS variants. Indeed, there is cautious security in maintaining a system that works adequately in its current configuration. The analog TCS performs well in its current form. It is certainly not because of performance deficits that an upgrade is being championed at the plant. So, a certain caution by the operators to embrace replacements may reside with the bias toward keeping the plant safe. A new TCS introduces some element of the unknown into operations. Only after the system has been fully validated and after operators have gained adequate hands-on training and experience can we expect accurate user satisfaction metrics. Of course, our comments are not an excuse to defend a poor HSI. Open-ended feedback from the operators expressed many positive impressions of the new TCS and very few concerns. It is therefore likely that the operators, while being positively disposed to the new TCS, are cautious to assign it high satisfaction until they have greater familiarity with the system.

4.3 Lessons Learned on Prototyping

The functional completeness of prototypes can influence user perception of the HSI. While the utility of low and medium fidelity prototypes has been amply demonstrated for consumer applications (e.g., Walker, Takayama, and Landay, 2002), the use of such prototypes for safety-critical applications may prove more challenging. The feedback received from the operators on the overview screen reflected to some extent a faulty indication in the prototype for the turbine turning gear. While this was a minor glitch in the prototype that was easily corrected, it proved a distraction from the ability of the operators to engage with or trust the system. Thus, several subjective scales used to measure operator satisfaction with the system reflected some minor dissatisfaction with the system relative to the TCS without the overview. At the same time, the overview afforded reduced time to complete tasks, as discussed in light of the performance-preference paradox (see Section 4.2). In this study, we compared two well-developed HSIs—the existing analog and the proposed digital TCS—with the overview prototype. The prototype for the overview screen represented early design concepts. The study demonstrated that mixing and matching designs at different stages of maturity may pose hindrances to operator acceptance of the designs. This may paint a troubling picture for the prospect of prototypes for safety-critical systems. A prototype for an NPP HSI may require a greater degree of fidelity than would be required in less exacting domains. Modernizing an HSI may risk pitting a mature system against a conceptual design, which may not always be a suitable comparison.

4.4 Lessons Learned on Performance Measures

The study investigated two new performance measures—the plant parameters and the performance drivers. Both showed promise as methods to expand the repertoire of measures available to evaluate the effectiveness of new HSIs. Because operator accuracy reaches asymptote due to the highly skilled and trained nature of NPP operations, unless deliberately challenging tasks are seeded into the scenarios (Boring, Lew, Ulrich, and Savchenko, 2016), conventional measures are likely not effective at benchmarking different types of technological systems such as comparing analog vs. digital control systems. Moreover, because some measures like workload and situation awareness may experience similar plateaus or prove insensitive to highly skilled operations, these mainstays of human factors evaluation may not be ideal candidates for validating new HSIs. We have made inroads on appreciating the value of qualitative insights, especially for informing formative design activities (Boring, 2015), but there still remains a need for quantitative metrics. This report has shown that additional measures can be derived from insights on performance drivers from human reliability analysis (Boring, Blackman, and Rasmussen, 2017). Additionally, we demonstrated that plant parameters can be used to show differences in performance of operators using the system. The jagged response curves of manually tuning turbine settings for load following with the analog TCS contrast with the much smoother curves found in the digital TCS for the same activity. Additional insights may be found by exploring different plant parameters, offering new ways to validate the design, not just by what the operator does or reports they do but also in terms of how the plant responds to operator actions, whether analog or digital, manual or automated.

4.5 Future Studies

While OSSO-2 was a successful comparison of three different TCS technologies, there is room to expand on the current study to answer methodological, scientific, and practical questions related to control room modernization. A few areas for future study include:

- *Comparison between component, system overview, and plant overview screens.* OSSO-1 and -2 have provided new insights into the use of system overview screens. It remains to be seen how system overviews compare to plant overviews. Data specifically comparing the three levels of screens will allow formalization of first-of-a-kind design guidelines on how operators go between sources of information and how this information use should be mapped to digital displays. It is hypothesized that the proper distribution of information between plant overview, system overview, and component control displays will minimize the need to navigate between nested windows in DCSs and increase situation awareness.
- *Development of a new measure of mutual awareness.* While there are many measures of individual and team situation awareness—the degree to which the operator and crews know what's happening at any given time—there are currently no good measures on mutual awareness—the degree to which the operators know what other operators are doing. There is evidence to suggest that digital HSIs may contribute to a breakdown in mutual awareness and threeway communication (Savchenko et al., 2017). A new measure is being developed to assess mutual awareness in the crew across scenarios and highlight any emerging degradations in mutual awareness. An operator-in-the-loop study continuing the OSSO lineage would be an idea testbed to validate the mutual awareness measure.
- *Color saliency of key indicators.* There are certain HSI design philosophies such as the dullscreen concept (Braseth, 2014) that prescribe a muted color palette. While these principles have driven the design of digital HSIs, they remain largely untested. Evidence from OSSO-1 (Lew, Ulrich, and Boring, 2017) suggests that operators found muted color palettes reduced their ability to pick out key parameter information (e.g., valve open vs. closed) at a glance or from across the control room. Thus, dullscreen may prove too dull for some types of information, and perhaps color should not be reserved only for alarm situations. An additional operator-in-the-loop study would allow INL staff to test color saliency with operators to develop better guidance on the use of color in digital HSIs.

- *Augmented displays with prognostic information.* It is possible to use prognostic systems (Vilim et al., 2013) to help operators detect faults in the system, from small leaks that would otherwise go undetected until manifesting larger consequences, to data disparities such as might be found in unintentional or malicious cyber faults. An operator-in-the-loop study allows demonstrating how to present such information optimally. A major concern is that operators may already experience information overload (e.g., alarm flooding) when there are system faults. Introducing additional early or probabilistic warning systems risks further overloading operators. An additional study would allow preliminary review of ways to present such information to operators that allows them to filter the information to prevent creating new forms of nuisance alarms.
- *Use of larger sample size.* There were a number of interesting comparisons found in the data presented in Chapter 3 of this report. In all but a few cases, there were no statistically significant differences between HSI conditions. This lack of significance may be due to a true lack of differences, but it may also be there is simply insufficient statistical power to reveal any differences. To improve statistical power, it is necessary to increase the sample size or increase the number of iterations for each condition using a repeated measures design. Repeated measures may prove difficult with full-task scenarios and a system like turbine control, because the system actually contains a minimal gamut of functions, making it impossible to parse out many nuanced and varied scenarios. The sample size can be increased by bringing in additional operators into the simulator. Since this is an expensive and time-consuming prospect, it may be possible to bring prototypes to the operators. A trade study conducted on DCS input devices (Ulrich, Boring, and Lew, 2015) demonstrates a viable method to bring a prototype system to the plant, evaluate a very specific scenario, and gain wider participation. In the trade study, 25 operators participated individually. Where crew performance is not of primary consideration, the sample size may be increased by running operators individually as participants. These considerations notwithstanding, it remains to be said that a perfect experimental design to glean statistical inference may not make the perfect study for formative evaluation. Insights from operators appropriate to improving the design rarely reside in the tableau of statistical findings; rather, they are often qualitative expressions of operator expertise. The study must not compromise operational realism or practical application in the quest for methodological purism or statistical significance. Good design for control rooms is the culmination of both qualitative operational insights and quantitative performance measures. Future studies should continue to strike this important balance.

5. References

Al Rashdan, A., Lew, R., Hanes, L., Kovesdi, C., Boring, R., Rice, B., and Ulrich, T. (2017). *The Operator Study on System Overviews (OSSO): Design Study for Digital Upgrades in Control Rooms with and without Overview Screens, INL/EXT-17-43423.* Idaho Falls: Idaho National Laboratory.

Boring, R.L. (2013). Lessons learned using a full-scale glasstop simulator for control room modernization in nuclear power plants. *Proceedings of the Human Factors and Ergonomics Society 57th Annual Meeting*, pp. 1712-1716.

Boring, R.L. (2014). Human factors design, verification, and validation for two types of control room upgrades at a nuclear power plant. *Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting*, pp. 2295-2299.

Boring, R.L. (2015). Envy in V&V: An opinion piece on new directions for verification and validation in nuclear power plants. *Proceedings of the 59th Annual Meeting of the Human Factors and Ergonomics Society,* 1746-1750.

Boring, R.L. (2016). As low as reasonable assessment (ALARA): Applying discount usability to control room verification and validation. *Risk, Reliability and Safety: Innovating Theory and Practice, Proceedings of the European Safety and Reliability Conference,* 950-955.

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.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., Blackman, H., and Rasmussen, M. (2017). Preliminary considerations for human reliability analysis for verification and validation of new designs. *PSAM Topical Conference on Human Reliability, Quantitative Human Factors, and Risk Management*, Munich, Germany.

Boring, R.L., and Joe, J.C. (2015). Baseline evaluations to support control room modernization at nuclear power plants. *9th International Topical Meeting on Nuclear Power Plant Instrumentation, Control, and Human-Machine Interface Technologies (NPIC&HMIT),* pp. 2155-2165.

Boring, R.L., Joe, J.C., Ulrich, T.A. and Lew, R. (2015). *Light Water Reactor Sustainability Program 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., and Lau, N. (2016). Measurement sufficiency versus completeness: Integrating safety cases into verification and validation in nuclear control room modernization*. Advances in Intelligent Systems and Computing*, *495*, 79-90.

Boring, R.L., Lew, R., & Ulrich, T.A. (2016). Epistemiation: An approach for knowledge elicitation of expert users during product design. *Proceedings of the 60th Annual Meeting of the Human Factors and Ergonomics Society*, 1699-1703.

Boring, R., Lew, R., and Ulrich, T. (2017). Advanced nuclear interface modeling environment (ANIME): A tool for developing human-computer interfaces for experimental process control systems. *Lecture Notes in Computer Science, 10293,* 3-15.

Boring, R., Lew, R., Ulrich, T., and Savchenko, K. (2016). When human error is good: Applications of beneficial error seeding. *13th International Conference on Probabilistic Safety Assessment and Management (PSAM 13),* Paper A-532, pp. 1- 7.

Boring, R.L., Ulrich, T.A., Joe, J.C., and Lew, R.T. (2015). Guideline for operational nuclear usability and knowledge elicitation (GONUKE). *Procedia Manufacturing, 3,* 1327-1334.

Boring, R., Ulrich, T., and Lew, R. (2016). RevealFlow: A process control visualization framework. *Lecture Notes in Artificial Intelligence*, *9744*, pp. 145-156.

Braseth, A. (2014). Information-rich design for large-screen displays. *Nuclear Engineering International, 59*, 22-24.

Carol, J.M. (2000). Five reasons for scenario-based design. *Interacting with Computers, 13,* 43-60.

Cooper, A. (2004). *The Inmates are Running the Asylum: Why High-Tech Products Drive Us Crazy and How to Restore the Sanity, Second Edition.* Indianapolis: Sams Publishing.

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience.* New York: Harper & Row.

Hart, S.G., and Staveland, L.E. (1988). "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research". In P.A. Hancock & N. Meshkati (Eds.), *Advances in Psychology, 52*.

Hildebrandt, M., and Fernandes, A. (2016). Micro task evaluation of analog vs. digital power plant control room interfaces. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 60,* 1349-1353.

Jokstad, H., and Boring, R. (2015). Bridging the gap: Adapting advanced display technologies for use in hybrid control rooms. *9th International Topical Meeting on Nuclear Power Plant Instrumentation, Control, and Human-Machine Interface Technologies (NPIC&HMIT),* pp. 535-544.

Kaufman, R., Rojas, A.M., and Mayer, H. (1993). *Needs Assessment: A User's Guide.* Englewood Cliffs, NJ: Educational Technology Publications.

Kolaczkowski, A., Forester, J., Lois, E., and Cooper, S. (2005). *Good Practices for Implementing Human Reliability Analysis, NUREG-1792.* Washington, DC: U.S. Nuclear Regulatory Commission.

Kovesdi, C., Spielman, Z., Le Blanc, K., and Rice, B. (2017). Application of eye tracking for measurement and evaluation in human factors studies in control room modernization. *Proceedings of the American Nuclear Society Nuclear Power Instrumentation, Control, and Human-Machine Interface Technology Conference,* pp. 1544-1556.

Lew, R., Ulrich, T.A., and Boring, R.L. (2017). Nuclear reactor crew evaluation of a computerized operator support system HMI for chemical and volume control system. *Lecture Notes in Artificial Intelligence, 10285,* 501-513.

McLeod, R.W. (2015). *Designing for Human Reliability: Human Factors Engineering in the Oil, Gas, and Process Industries.* Amsterdam: Elsevier.

Nielsen, J. (January 12, 2014). *Turn User Goals into Task Scenarios for Usability Testing.* Article retrieved from https://www.nngroup.com/articles/task-scenarios-usability-testing/

O'Hara, J.M., Higgins, J.C., Fleger, S.A., and Pieringer, P.A. (2012). *Human Factors Engineering Program Review Model, NUREG-0711, Rev. 3.* Washington, DC: U.S. Nuclear Regulatory Commission.

Rasmussen, M., Boring, R., Ulrich, T., and Ewing, S. (2017). The virtual human reliability analyst. *Advances in Intelligent Systems and Computing, 589,* 250-260.

Savchenko, K., Medema, H., Boring, R., and Ulrich, T. (2017). Comparison of mutual awareness in analog vs. digital control rooms. *Advances in Intelligent Systems and Computing, 589,* 192-199.

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 (SEE N 90-28972 23-53)*.

Ulrich, T.A., Boring, R.L., and Lew, R. (2015). Control board digital interface input devices—Touchscreen, trackpad, or mouse? *Proceedings of the International Symposium on Resilient Control Systems (Resilience Week),* pp. 168-173.

Ulrich, T.A., Lew, R., Boring, R.L., Thomas, K.D., Rice, B., and Poresky, C.M. (2017). *Operator-in-the-Loop Study for a Computerized Operator Support System (COSS)—Cross-System and System Independent Evaluations, INL/EXT-17-43390.* Idaho Falls: Idaho National Laboratory.

U.S. Nuclear Regulatory Commission. (1981). *Human Factors Acceptance Criteria for the Safety Parameter Display System, NUREG-0835.* Washington, DC: U.S. Nuclear Regulatory Commission.

Vilim, R.B., Park, Y.S., Heifetz, A., Pu, W., Passerini, S., and Grelle, A. (2013). Monitoring and diagnosis of equipment faults. *Nuclear Engineering International, 58,* 24-27.

Walker, M., Takayam, L., and Landay, J.A. (2002). High-fidelity or low-fidelity, paper or computer? Choosing attributes when testing web prototypes. *Proceedings and the Human Factors and Ergonomics Society Annual Meeting,* pp. 661-665.

Woods, D. (1995). Toward a theoretical base for representation design in the computer medium: Ecological perception and aiding human cognition. In J. Flach, P. Hancock, J. Caird, and K. Vicente (Eds.), *Global Perspectives on the Ecology of Human-Machine Systems.* Hillsdale, NJ: Lawrence Erlbaum Associates.