INL/CON-16-38083
PREPRINT

Full Scale Evaluation of How Task-Based Overview Displays Impact Operator Workload and Situation Awareness When in Emergency Procedure Space

7th International Conference on Applied Human Factors and Ergonomics (2016)

Zachary Spielman, Rachael Hill,
Katya LeBlanc, Brandon Rice, Gordon Bower,
Jeffrey Joe, David Powers

July 2016

Full Scale Evaluation of How Task-Based Overview Displays Impact Operator Workload and Situation Awareness When in Emergency Procedure Space

Zachary Spielman, Rachael Hill, Katya LeBlanc, Brandon Rice, Gordon Bower, Jeffrey Joe, and David Powers

Idaho National Laboratory,
2525 Fremont Avenue, PO Box 1625, Idaho Falls, Idaho 83415-3818
{zachary.spielman, rachael.hill, katya.leblanc, brandon.rice, gordon.bower, jeffrey.joe@inl.gov}

Abstract. Control room modernization is critical to extending the life of the 99 operating commercial nuclear power plants (NPP) within the United States. However, due to the lack of evidence demonstrating the efficiency and effectiveness of recent candidate technologies, current NPP control rooms operate without the benefit of various newer technologies now available. As nuclear power plants begin to extend their licenses to continue operating for another 20 years, there is increased interest in modernizing the control room and supplementing the existing control boards with advanced technologies. As part of a series of studies investigating the benefits of advanced control room technologies, the researchers conducted an experimental study to observe the effect of Task-Based Overview Displays (TODs) on operator workload and situation awareness (SA) while completing typical operating scenarios. Researchers employed the Situation Awareness Rating Technique (SART) and the NASA Task Load Index (TLX) as construct measures.

Keywords: Nuclear Power Plant, Control Room, HSI, Automation, Human-Automation Collaboration, Human Performance, Situation Awareness, overview displays, workload

1 Introduction

Background. The following research is part of the United States (U.S.) Department of Energy (DOE) sponsored Light Water Reactor Sustainability (LWRS) Program conducted at the Idaho National Laboratory (INL). The LWRS program is a collaborative effort with industry research and development (R&D) programs to establish the technical foundations for licensing and managing the long term, safety, and economical operation of current nuclear power plants (NPPs). One primary mission of the LWRS program is helping the U.S. nuclear industry modernize operating capabilities by adopting tested technologies and engineering solutions found to facilitate safe operation through a NPP's extended lifetime (1). The Electrical Power Research Institute (EPRI) describes several potential drivers of control room modernization including [2]:

1. To address obsolescence and lack of spare parts
2. To meet the need for equipment replacement due to high maintenance cost or lack of vendor support for existing equipment
3. To implement new functionality necessary for adding beneficial capabilities
4. To improve plant performance, HSI functionality, and reliability
5. To enhance operator performance and reliability
6. To address the difficulties in finding young professionals with education and experience with older analog technology

Items one and two both refer to the growing need that must be addressed for a NPP to continue operating while items three through six address possible improvements that should be considered during a modernization.

Despite the six drivers listed; none of the 99 currently operating NPPs within the US have undergone a full control room modernization [3]. Current control rooms are not built to facilitate a simple modernization effort. Between various regulatory, operations benchmarks, and licensing hurdles the control room requires careful deliberation as to what updates occur first. Note also that any physical changes to the current analog boards would be, for all practical purposes, permanent. Due to such difficulty a hesitation towards modernization exists because the evidence demonstrating the value of candidate technologies that goes beyond the original design concept (i.e. more than replacing analog alarm panels with identical panels) is lacking. Therefore strong evidence to validate the struggle and work required to modernize must be found before a plant can commit to upgrading or supplementing the control room. As a result current NPP control rooms operate without the benefit of various capabilities newer technologies offer.

However, as NPPs have been extending their licenses to continue operating for another 20 years the interest towards modernizing the control room and supplementing the existing control boards with advanced technologies has increased. The recent boost of interest has provided the opportunity to work with the industry to identify technology that improves upon current operating systems. The LWRS project aims to provide industry with tested solutions expected to exceed the performance of the current labor intensive model and address the latter items on EPRI's list.

Hybrid Approach. Hybrid control rooms with supplemental technology have been proposed as an initial solution to the complicated modernization process. As a stepping stone to full modernization, new technologies might supplement rather than replace portions of the control room [3]. Furthermore plants can make changes to the control room in a shorter amount of time, minimize permanent changes to the analog boards, and is generally less of an investment. Supplementing offers NPPs the opportunity to 'test drive' the technology as a potential replacement before committing to a physical redesign of the control room. To minimize the number of failed 'test drives', candidate technologies undergo testing to guide decisions towards the best options. However, the attempt to evaluate candidate technologies within a full-scale main control room operating context is accompanied by challenges in recreating realistic scenarios and testing real world operators. The challenges include:

- Finding available and experienced crews of operators
- Comparing candidate technology using unbiased expert operators
- Training operators on a generic operating system within a feasible time span

- Developing a facility flexible enough to mimic various NPP control rooms
- Measuring impact of candidate technology on operator-system interaction

Due to the complexity of control room operations, real-world operators in a high-fidelity control room will be used to test candidate technologies in a control room setting to gain results generalizable to the industry as a whole. However, using real operators in a familiar control room may bring in operator bias towards the familiar. Instrumental in overcoming these challenges is the Human-System Simulation Laboratory (HSSL). The lab hosts a simulated operator control room environment equipped for testing candidate technologies. The flexible interface can simulate either a generic pressurized water reactor (gPWR) or the control boards of an industry control room depending on need. The lab is equipped with various recording technology for data collection and, later, analysis.

Recently a pilot study was launched to begin addressing these challenges and develop a protocol to achieve meaningful results without sacrificing fidelity. Researchers under the LWRS pathway at INL are investigating which near-term control room technologies may enhance operator performance in hybrid control rooms.

The advantage to using the gPWR is to reduce bias a simulated by creating even ground for all operators; no single operator will have more experience in the generic control room than another. The generic control room also supports the ability to generalize results across various plants. Once potential candidate technology is identified using the gPWR simulation, the HSSL can mimic the control rooms of industry partners interested in seeing how the technology affects their particular plant operations method.

Finding which candidate technologies improve operator performance is expected to lead to updating current control room technology as well as inform design concepts of future control rooms and continue evaluation of evolving or novel candidate technologies. The results described here are the beginning of an ongoing effort to verify and validate the impact candidate technologies have on operator and plant performance.

Researchers at the INL conducted a full-scale evaluation and began establishing a method of testing sensitive to changes in performance caused by candidate technologies. The study observed the effect of Task-based Overview Displays (TODs) on operator workload and SA while completing typical operating scenarios.

Task-based Overview Displays. The TODs are expected to reduce operator workload and improve SA by using information-rich design principles to provide many variables and trends occurring in a NPP simultaneously at a given moment. The displays are “Task based” because the available content is dependent on current plant state. Content important for monitoring and maintaining a plant in normal operating procedure space is available during regular procedures. When the plant enters emergency procedure space post reactor trip, the available information is adapted to provide the crew with information relevant to the tasks involved in emergency procedure space [4].

Using trend displays and graphical representations of pumps, valves, and flow balance the TODs break from traditional analog displays to provide greater amounts of information with a single display element to increase the rate and ease of information gathering to improve situation awareness when a plant condition changes. Currently, most NPP control rooms are comprised of "boards" that contain the controls, alarms, and indicators, to specific aspects of the plant. Normally, operators must monitor a range of dials and indicators that requires them to move up and down the boards, or 'ping-pong' between boards to retrieve the information necessary to successfully complete their task. The TODs are intended to reduce the need to ping-pong by placing task relevant information for a single operator in a single location easily accessible from any position along the boards. For a detailed description of the design principles used in the TOD's refer to [4].

2 Methods

2.1 Participants

INL employees with nuclear operations experience were recruited to participate in this study. Seven operators participated serving the roles of Senior Reactor Operator (SRO), Reactor Operator (RO) and Balance-Of-Plant Operator (BOP). All the participants were male averaging 48.6 years of age. All of the participants have bachelor's degrees in nuclear/mechanical engineering and experience as Navy nuclear operators; however none of the participants currently work in operations.

2.2 Environment & Stimuli

The HSSL houses many integral components used to complete this study. Most notably is the full scale, full scope, and reconfigurable virtual NPP control room simulator. The simulator consists of fifteen bays each consisting of three, 47 inch LCD screens (measured diagonally). The bottom two LCDs have touch-screen capabilities via infrared overlays. A Dell OptiPlex desktop computer running Microsoft Windows 7 Professional is housed inside each of the bays, and acts as the client to the simulator software code running on a secure server. The server room houses backend servers that allow for rapid image deployment via Free Open Ghost (FOG), Windows Server 2008 R2 for different plant models and configurations, and Microsoft Hyper-V utilization to satisfy virtualization needs. The bays are mounted on frames with lockable wheels for mobility, maintainability and convenience. Because of these features, the control room is reconfigurable into almost any NPP control room layout.

Other resources used in the HSSL include virtual machines, an air-gapped network infrastructure, Foscam wireless Internet Protocol cameras for video capture, and Peavey wireless lavaliere microphones for audio capture. Blue Iris software is used to record and synchronize the audio and video feeds. There is also an observation room in the HSSL, in which resides a Dell OptiPlex computer that serves as the Instructor Station for the simulator. From there, all simulator activities are controlled, including: powering on the bays, starting the simulator, loading initial conditions, inserting a malfunction scenario, and powering down the bays. The core of the system is run by GSE Systems Java Application Development Environment (JADE) simulator platform.

Figure 1: The Human System Simulation Laboratory (HSSL) with the generic pressurized water reactor control boards displayed on the three-screen bays.

2.3 Scenarios

The scenarios were designed as two sets: each set with relative similarity in complexity and difficulty. Furthermore, scenarios were designed to move operators from normal procedure space to emergency procedure space. The movement through these procedure spaces is important as there are many decision gates the crew must accurately navigate. Decisions are easy targets when evaluating crew performance. Either the crew makes a correct or incorrect decision at each point providing a simple evaluation for non-expert observers. Scenario set B was the practice set used to familiarize the operating crews with the plant. The set of scenarios consisted of:

A. Scenario Set A: Loss of Coolant Accident (LOCA)
   1. Simple Case
   2. Complex Case
B. Scenario Set B: Faulted Steam Generator (SG)
   1. Simple Case
   2. Complex Case
C. Scenario Set C: Steam Generator Tube Rupture (SGTR)
   1. Simple Case
   2. Complex Case

A1: Simple LOCA. This scenario is a standard LOCA, a commonly trained scenario in commercial NPPs. The only complication was the leak occurring during a power ramp, meaning the crew will already be involved in a procedure as the break begins. The scenario progresses relatively slowly, giving the crew ample time to react and identify the issue before any critical issues occur. The scenario is ended once the affected leg is isolated and the crew is transitioning to "post-LOCA cooldown and depressurization".

A2: Complex LOCA. Scenario A2 progresses quickly and has a masking fault making it the more complex of the two LOCA scenarios. Heater drain pumps are programmed to fail sequentially within a minute of each other causing alarms to trip and the crew to take action. As the crew is working to resolve the heater drain pump issue a large LOCA is inserted immediately tripping the reactor. Crews have to work quickly to stabilize the reactor and diagnose the issue. The scenario will end as the crew begins establishing alternate cooling routes.

B1: Simple Faulted Steam Generator. The scenario begins with a masking fault when charging pump 'A' experiences a shaft shear rendering the pump ineffective however indicators will not pick up the pump failing, only the flow ceasing. As the crews attempt to resolve the situation a steam generator experiences a main line break outside of containment. The resulting action moves the crew to emergency procedure space where they continue to stabilize the plant and diagnose the issue.

B2: Complex Faulted Steam Generator. The scenario begins with a masking fault when a charging line begins to slowly leak at 50 gallons per minute (gpm) moving the crew into abnormal procedure space. As the crew is determining if they can maintain the leak and still remain in operation the main steam line break is inserted automatically tripping the reactor. The reactor trip, as always, moves the crew into emergency procedure space. During this reactor trip however, the turbine does not automatically trip with the reactor as it is designed to do. The crew is monitored to be sure they detect the abnormality in system operation. The scenario is ended once the crew reestablishes continued Reactor Coolant System flow.

C1: Simple SGTR. The scenario begins with a masking fault, the failure of a pressurizer instrument expected to be identified before the SGTR begins and does not significantly affect diagnosis. The pressurizer indicator shows a sudden drop to the low end, causing alarms as well as three heater pumps to turn on as the system attempts to increase pressure. A redundant indicator that does not break should indicate to the crew that the indicator is malfunctioning. The crew is expected to appropriately handle the malfunction and at the resolution of which an SGTR in is injected. A slowly developing leak eventually leads to charging flow and pressurizer alarms. The size of the leak will exceed makeup capabilities and the crew will need to trip the reactor (an automatic trip would represent a failure condition). The scenario concludes when the crew has correctly identified the issue.

C2: Complex SGTR. The scenario begins normally with a masking fault, a tripped heat exchanger pump, inserted shortly after. The standby pump will start from low system pressure almost immediately, and the crew will progress through Abnormal Operation Procedure space (AOP) to confirm that this has happened and that parameters are returning to normal. As the crew is working through AOP, the SGTR is inserted. The team should eventually transition into the appropriate AOP once loss of primary side coolant is identified. This should progress similarly to the previous SGTR, with step 4 as a possible trip point depending on how quickly the crew has progressed but eventually the reactor will trip. The scenario concludes when the crew has correctly identified the issue.

2.4 Measures

Identifying differences in situation awareness and workload afforded by the TODs was the research objective. By counterbalancing which scenarios were accompanied by the TODs between crews and standardizing the questionnaires across conditions the researchers attempted to isolate how the displays impacted the two aspects of operator working space. The following describes the measures used and the experiment design.

Researchers used the Situation Awareness Rating Technique (SART) to determine the level of participant's situation awareness with and without the OVDs. Often utilized for aircrew studies, SART is a common questionnaire validated for measuring SA for the duration of a task [5]. The questionnaire is administered after a completed trial and presented in an ordinal format with nine items rated by the participants on a scale of one to seven (low to high).

- Instability of situation
- Variability of situation
- Complexity of situation
- Arousal
- Concentration of attention
- Division of attention
- Spare mental capacity
- Information quantity
- Familiarity with situation

The NASA-TLX was used to measure the subjective workload of the participants after completing a scenario. This evaluation is a subjective six-item scale that is a widely used and validated scale for measuring workload after a task. It was developed specifically for the aviation industry [6], though it has been used in many studies in a wide variety of fields, including many NPP control room studies [7]. It is accepted as a reliable measure of workload differences between tasks. It measures six constructs on a scale of one to ten.

- Frustration Level
- Effort
- Performance
- Temporal demand
- Physical demand
- Mental demand

The posttest questionnaire was used to gather additional data not already captured by NASA-TLX and SART. It includes four follow up items;

- How difficult was the scenario?
- How was your performance during the scenario?
- Were the overview displays available during the scenario?
- If the overview displays were available during the scenario, did you use them? Why or why not?

There were additional measures that were recorded but were not integral to the purposes of this report.

2.5 Procedure

Participants with a background in nuclear reactor operations were recruited for the study. They first provided informed consent to participate in the study acknowledging that they understood their rights during the study. Each crew worked together over three contiguous days to complete training and testing together. A day was considered an eight hour business day.

Day one began with training videos to familiarize the participants with the mechanical structure of gPWR and how to operate the simulated plant using a combination of video and referencing piping and instrumentation diagrams. The crews were allowed to ask questions and find what was being described on the video, on the simulator boards.

Day two continued the mechanical over view as well as the conduct of operations for the plant. Participants were given intermittent opportunities to “walk the boards” and interact with the simulator performing common actions associated with their role in the plant operations. Day two concluded with operators completing two formal training scenarios; a simple and a complex ‘steam-generator tube rupture’.

Day three consisted of four testing scenarios used to evaluate how the candidate technologies impacted operator performance. The four scenarios were constructed in a 4x2 fashion. Two types of scenarios of two complexity levels were ran; a “Loss of Coolant Accident” and a “Steam Generator Failure”. Each type came in two versions, with and without the TODs present. The order was counterbalanced across crews to account for learning effects.

Protocol was the same across every scenario. Crews were both video and audio recorded during all scenarios. The participants were handed over operation of the plant at full power with no known or occurring issue with exception to a single scenario requiring a power ramp. Once in control, the crews monitored the plant as faults were entered from the observation room by the simulator controller located behind the operators. All Scenario details were guarded from the crew beforehand to gain as naturalistic responses as possible. Operators were also not aware of which scenarios the TODs would be available. However the researchers had established a display order a priori. When the displays were available it was clear to the operators.

During the scenario crews were stopped at three different points to fill out a short freeze-probe questionnaire requiring only couple minutes at which point all simulator screens were blacked out to remove any hints or answers to the questions. The crews were allowed a quick brief to refocus on the task at hand right before resuming the simulation. At the conclusion of each scenario, each crew member filled out the NASA-TLX, the SART, and the post-test questionnaire. Note the post-test questionnaire asked if displays were present during a scenario to ensure the crews were aware of the TOD’s provided them. During this time the simulator was reset, data exported, and procedure lists refreshed.

3 Results

Overall crews were able to carry out scenarios to the pre-selected termination point. Crews were also able to answer the questionnaires easily with minimal questions themselves. The final day of testing went smooth for all crews with a small adjustment period for the final crew as one member cancelled last minute.

Using the Likert Scale values collected in the SART and NASA TLX, researchers made many comparisons to determine where meaningful differences occurred in the study. All the individual questions from the self-reporting measures were kept separated to search for differences within the questionnaires (i.e. in mental demand but not physical demand of the NASA TLX). Then researchers made pair-wise comparisons across three different pairs. Each pair was carefully matched to have only a single difference in treatment. The following pairs were compared across both measures:

1. TOD on/off condition broken down by crew designation
   - RO
   - SRO
   - BOP
2. Scenario complexity
   - Simple
   - Complex
3. Scenario complexity by crew designation
   - RO
     - Simple
     - Complex
   - SRO
     - Simple
     - Complex
   - BOP
     - Simple
     - Complex
4. Complexity by Task Display
   - Simple
     - On
     - Off
   - Complex
     - On
     - Off

The first comparison was intended to determine the candidate technology's impact on operator performance. The different crew designations have different tasks and different overview displays to use. Breaking down TODs by crew designation was a natural decision to understand the technology's impact on individual roles in the control room. The results found no difference between having and not having the displays present during the scenario for any of the three roles.

The second comparison was important towards determining if the different scenario complexities had an impact on operator workload to verify their intent. However, again no significant differences could be found to determine if simple scenarios were easier to complete than complex scenarios. The third comparison took a deeper look to see how scenario complexity affected individual roles. The result was no different than the second comparison.

The fourth comparison was used to evaluate how the presence of theoretically more accessible information made a difference when the need for information was increased. No difference was discovered within the fourth comparison either. Since a difference between the complex and simple scenarios was not confirmed this does not come as a surprise.

The post-test debriefing asked for operators subjective input regarding the use of the displays. Of 14 responses total one said the displays were not useful attributing their remark to the number of valve changes they were doing. The positive statements often contained similar content and more than one reason stating the TODs were helpful. The positive 13 responses included statements similar to the following:

- "Yes, useful for tracking all trends…" (4/13)
- "Ability to validate plant information" (4/13)
- "Yes, good indicators for checking [Steam Generators]" (3/13)
- "Checked pressurizer display" (5/13)

Thus the operator's majority responses were positive towards the presence of the TOD's.

4 Discussion

The researchers evaluated whether an environment was able to test crew performance while manipulating the various candidate technologies and observing the resulting effect on crew performance, SA, and workload. It was understood from the beginning that not every measure may produce a measurable result due to the small sample size. Such a result was acceptable as another goal was testing the facilities ability to collect measures, incorporate new technologies, and host operating crews able to perform scenarios at an acceptable level.

The results of the measures were non-descript as to whether the TODs made an impact on operator performance. It appears as though there was no difference made however with such a small sample size it is far more likely the measures were not sensitive to changes in performance. The finding exemplifies the need for multiple crews to perform the tasks in order to validate the testing measures and scenario differences. Due to the positive responses from all operators regarding the TOD's there is promise the displays may have a positive impact on performance.

The project was successful in addressing challenges associated with evaluating candidate technologies within a full-scale main control room operating context. While some limitations were present in reducing the number of crews the project could host, a viable source at the INL was found in previously experienced Navy Nuclear operators. Additionally, these participants were able to train and carry out tasks on the gPWR within a feasible time limit. Further attention towards the training regime would likely increase the positive transfer and reduce the time required for training allowing more time for testing. Using ex-operators in a generic control room leads us to believe there is reduced bias towards using the familiar systems but without sufficient sample size it is difficult to determine exactly.

The HSSL demonstrated a functional flexibility that provided researchers with the necessary tools to administer various performance measures collected during this study. The observation room allows for unobstructed visual access to all the actions being taken in the control room while all plant parameters are recorded and loaded to a safe directory for later viewing. The simulator demonstrated the capability to freeze, hide all screens, and quickly unfreeze again during a scenario for freeze probe questionnaires. The crews could operate the simulator with enough proficiency to complete the tasks without their inability hindering their task completion.

Overall, the study has demonstrated the HSSL has all the capability to perform full-scale evaluations of crew performance during control room scenarios. The laboratory has the flexibility to work with varying candidate technologies in an applied workspace. Furthermore multiple measures may be gathered at once and stored in a single place. The training developed for the study was sufficient in training operators to carry out scenarios on the generic plant in a feasible time period. Finally, to gain meaningful results the number of crews run and the resources to do so need to be considered when beginning such an endeavor to gain the most information from a single evaluation. The next task is to bring in a sufficient number of experienced operating crews and begin making decisions towards how the control room can best be upgraded.

References

1. United States of America, Department of Energy, Nuclear Energy. (2015). Light Water Reactor Sustainability Program Integrated Program Plan (Revision 3 ed., Vol. INL/EXT-11-23452, pp. 43-61).
2. Naser, J., Hanes, L., O'Hara, J., Fink, R., Hill, D., & Morris, G. (2004). Guidelines for Control Room Modernization as Part of Instrument and Control Modernization Programs.
3. LeBlanc, K., Joe, J., Rice, B., Ulrich, T., & Boring, R. (2015). Benefits of Advanced Control Room Technologies: Phase One Upgrades to the HSSL and Performance Measures, INL/EXT-15-3538 Revision 0.
4. Braseth, A. O. (2015). Information-Rich Design: A Concept for Large-Screen Display Graphics: Design Principles and Graphic Elements for Real-World Complex Processes.
5. Taylor, R. M. (1990). Situational Awareness Rating Technique (SART): The development of a tool for aircrew systems design. AGARD, Situational Awareness in Aerospace Operations 17 p(SEE N 90-28972 23-53).
6. Hart, S.G., 2006, NASA-task load index (NASA-TLX); 20 years later, In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Volume 50, No. 9, pp. 904-908), Sage Publications.
7. LeBlanc, K., Powers, D., Joe, J., Spielman, Z., Rice, B., & Fitzgerald, K. (2015). A Pilot Study Investigating the Effects of Advanced Nuclear Power Plant Control Room Technologies: Methods and Qualitative Results, INL/EXT-15-36432 Revision 0.