**INL/EXT-17-43250**
**Revision 0**

# Light Water Reactor Sustainability Program

# Evaluation of Control Room Interface Designs to Support Modernization in Nuclear Power Plants

September 2017

U.S. Department of Energy

Office of Nuclear Energy

INL/EXT-17-43250
Revision 0

# Evaluation of Control Room Interface Designs to Support Modernization in Nuclear Power Plants

Katya Le Blanc
Casey Kovesdi
Rachael Hill
Zachary Spielman
Tyson Hansen
Johanna Oxstrand

September 2017

Prepared for the
U.S. Department of Energy
Office of Nuclear Energy

# ABSTRACT

Control room modernization in the commercial nuclear power industry often takes the form of phased system-by-system upgrades that result in hybrid control rooms with digital and analog equipment. Ensuring that these upgrades result in enhanced human performance and more intuitive interfaces, researchers at INL have been supporting the design of upgraded systems in coordination with a large scale modernization effort at a partner utility. This report describes the experimental evaluation of a liquid radiological waste control room concept which investigated the use of color in system overview displays. The results indicate that maintaining existing color conventions to identify component status yielded no benefit, and in some cases caused a detriment to performance. The researchers recommended using a minimal muted color in overview displays, but advocated using color to enhance understanding of different flow paths and using saturated color to highlight abnormal operating states.

## ACKNOWLEDGEMENTS

We are grateful for the ongoing support from the radiological waste control room operators and engineering staff at Arizona Public Services. Further, we would like to thank the staff at the Institute of Energy Technology (IFE) who were an integral part in the experimental design for the operator workshop by providing technical expertise and support with the micro-task data collection activities. This report was made possible through funding by the U.S. Department of Energy (DOE) Light Water Reactor Sustainability Program. Lastly, we would like to thank Richard Reister of DOE and Ken Thomas and Bruce Hallbert of Idaho National Laboratory for championing this effort.

# CONTENTS



# FIGURES

## TABLES

# ACRONYMS

| | |
|---|---|
| AOI | Areas of Interest |
| BAC | Boric Acid Concentrator |
| DOE | Department of Energy |
| FC | Fixation Count |
| FD | Fixation Duration |
| HSI | Human-System Interface |
| INL | Idaho National Laboratory |
| LRS | Liquid Radiological Waste System |
| LTLFTR | Latency between Time to First Fixation to Correct Response |
| RT | Response Time |
| TTFF | Time to First Fixation |
| TTLF | Time to Last Fixation |
| XML | Extensible Markup Language |

# Evaluation of Control Room Interface Designs to Support Modernization in Nuclear Power Plants

## 1. Introduction

Throughout the past two decades, the nuclear industry has upgraded and modernized plant and control room systems to manage the impact of aging systems, structures, and components. Each upgrade involves an increased amount of advanced technology. This addition of advanced technology brings new opportunities for efficiency and safety improvements. However, if not carefully designed and planned, the incorporation of advanced technology could cause unintended negative consequences on both system and human performance.

Upgrades to main control rooms have been the subject of years of research (Boring, Agarwal, Joe, and Persensky 2012; Boring, Joe, Ulrich, and Lew 2014; Joe, Boring, and Persensky 2012; Ulrich, Boring, and Lew 2014). The main control room operators monitor and manage the major plant systems, such as the reactor and turbine systems. Due to the high cost and risk associated with full control room modernizations (i.e., a complete upgrade of all systems and panels in the main control room), very few have been conducted throughout the industry. Instead, most main control rooms today are hybrid control rooms, meaning both new and original systems exist together in a single control room.

Though undoubtedly important, the main control room is not the only control room in the nuclear power plant. In addition to the main control room, the plant usually has one or two smaller control rooms spread throughout the plant. These local control rooms tend to not be as safety critical as the main control room and are not as highly prioritized when planning future upgrades. Because of this lack of prioritization over time, these local control rooms may be in greater need of upgrades than the main control room. An example of such local control room is the radiological waste control room.

Human factors researchers at Idaho National Laboratory (INL) have been supporting control room modernization efforts in the nuclear industry for several years. The researchers have defined processes and design concepts to enhance human performance when legacy systems are replaced with new technology. The research has mainly focused on human factors issues related to main control room operations (Boring, Agarwal, Joe, and Persensky 2012; Boring, Joe, Ulrich, and Lew 2014; Joe, Boring, and Persensky 2012; Ulrich, Boring, and Lew 2014). However, starting in fiscal year 2017, as a part of a control room modification project in Department of Energy's (DOE's) Light Water Reactor Sustainability Program, the INL researchers are now collaborating with a nuclear utility to develop design concepts for their local radiological waste control rooms.

The radiological waste control room is not as strictly regulated as the main control room, which makes it a great testbed for new human factors design concepts. Hence, the researchers have the opportunity to design and evaluate innovative design concepts for the radiological waste control room. These concepts, if proven safe and effective, may not only impact the recommendations for human-system interfaces (HSI) for the radiological waste control room, but may also inform recommendations for the design of the main control rooms.

This report summarizes the activities conducted to meet the milestone M3LW-17IN06031414 "Complete Report on Control Room Technology Benefits Studies" This report describes the experimental results from a dynamic workshop to evaluate the design of an interface for the radiological waste control room. The report is structured as follows:

- Section 1.1 Describes the LRS control room
- Section 1.2 Describes the prototype interface and identifies unanswered design questions that will be addressed by experimental studies

- Section 2 Describes how the research addresses the design questions
- Section 2.1 describes the experimental methodology that investigates the design questions

## 1.1 Radiological Waste Control Room

The basic purpose of the radiological waste control room is to capture, store, and repurpose radiological waste. For example, through evaporation, both pure water and concentrated boric acid can be reclaimed from liquid radiological waste, both of which are essential for the operation of the plant. Radiological waste can come in solid, gas, or liquid form.

The actions performed to operate the radiological waste systems take place either in the local control room or out in the plant (i.e., cranking a hand wheel to open a valve). The radiological waste control room consists of two rows of large cabinets with controls and indicators (e.g., nobs, buttons, and meters), as illustrated in Figure 1. The utility's vision is that the cabinets will be replaced with a digital control system and that the operators will use a new digital human system interface on workstation monitors to operate the systems.

Figure 1. 3-D model of the existing radiological control room.

The research effort studies the three subsystems of the liquid radiological waste system (LRS), which are the liquid radiological waste evaporator, the boric acid concentrator (BAC), and the liquid radiological waste panel.

At the collaborating utility, there are only a handful radiological waste operators. Usually, one to two operators are stationed in the control room a couple times per year. The BAC is exercised three to four times per 18-month refueling cycle and the evaporator is used three times per year on average. It takes about 4 hours to start up the evaporator and the BAC takes about 1.5 hours to start up, if every component works as intended.

The evaporator's function is to remove water from and increase concentration of total solids in liquid waste. The function of the BAC is to process borated primary wastes from the chemical volume and control system hold up tank. The liquid radiological waste panel consists of five subsystems, including total dissolved solids, evaporator, concentrate monitor, recycle monitor, and chemical drain. The evaporator subsystem and the BAC perform the same functions, but for different purposes. Both systems evaporate water from the radiological waste; however, the evaporator reclaims and repurposes the clean water while the BAC system reclaims and repurposes the boric acid. Since these systems are so similar, the evaporator can be used as a backup to the BAC system if the BAC becomes unavailable.

The researchers are developing design concepts for the graphical user interface to be used by the operators to operate the LRS. This paper will describe the effort related to the evaporator subsystem.

## 1.2 Design Concept for the LRS Control Room

To facilitate development of HSI design concepts and evaluate those designs, the team developed an overall display philosophy for the systems in the radiological waste control room for the operator workstation concept. The concept includes system overview displays with embedded control for the evaporator and the boric acid concentrator. A third display will contain overview information for the rest of the liquid radiological waste system. The fourth display will contain alarm information. Each display will be presented on one of the four monitors in the control room.

Le Blanc et al., (Le Blanc, Boring, Joe, Hallbert, & Thomas, 2014) describes several features that are commonly found in overview displays. The team selected the characteristics that were incorporated into the radiological waste control room overview displays based on the benefits proposed and the constraints and opportunities provided by the radiological waste control room system itself.

The systems contained in the radiological waste control room are relatively simple compared to the overall plant systems. Further, even after the planned upgrades, many of the components will be operated locally. Therefore, it is possible to represent the portions of the system that will be operated from the control room on a single overview display. This allowed for the available task-related information and the relevant system information to be displayed on a single display.

The design of the overview display features a simplified mimic of the entire evaporator system with all of the remotely operated equipment represented. The data for the system is embedded within the system in mini-trends and micro-trends. The mini-trends present data for the controllers (i.e., flow and level controllers in the system) and contain alarm set point information as well as trending of the control parameters. The set points are also represented on the mini-trend. All other system parameters are shown on micro-trends, which show the digital value of the parameter at its current level and a brief historical trend of the parameter.

Control is embedded on the overview display. The operator must simply click on a component (e.g., valve or pump) to bring up the associated controller faceplate. The faceplates pop up in a dedicated portion of the screen on the bottom-right side. The components of an active control faceplate are highlighted with a blue halo. For controllers, the operator can click on a small button to the left of the mini-trend to bring up the controller. Clicking on any of the components associated with controller will also bring up the controller (i.e., clicking on a level valve will bring up the level controller).

The researchers adopted a semi-dull screen approach to color. The majority of the static elements in the screen are presented in shades of gray. Dynamic data is presented in a green (a standard color used to represent live data) and active components are highlighted in white (they are gray when closed or off). The only static portions of the screen that are colored are the different flow paths for the product streams, and these are presented in muted colors. Dynamic information such as highlighting the selected components or the status of components is presented in a bright blue. Alarm states are presented in a saturated red. Figure 2 shows the overview and illustrates the concepts discussed above.

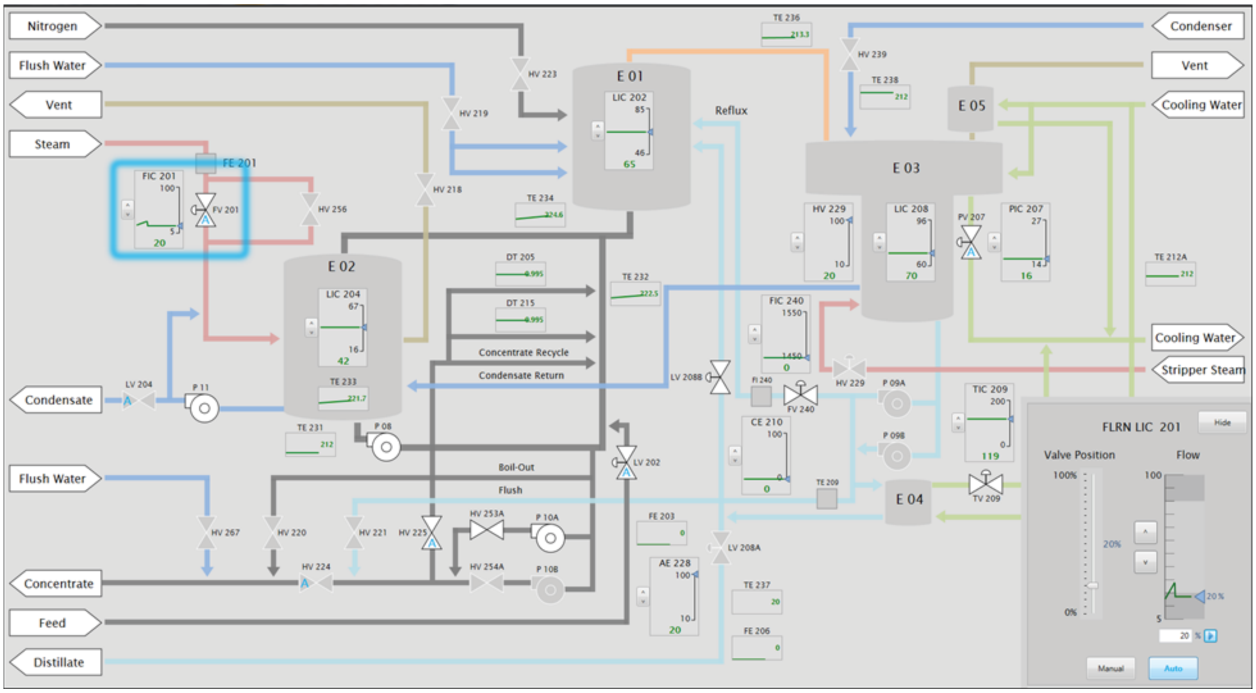



Figure 2. Evaporator system overview with embedded control.

One of the objectives of the design workshop was to empirically investigate some of the design questions early in the design phase so that decisions have a sound basis. One of the main questions at this stage is whether to use a dull-screen concept, to use a more colorful screen with typical red/green color scheme to represent component status, or to use some compromise between the two. Dull screens have the advantage of making information like alarms and dynamic data more salient than the static elements of the screen, which may potentially enhance detection of important information. Generally, operators prefer to have the display preserve the red/green stereotypes for valve position and pump status, because they argue that it helps them identify valve alignment more easily. However, this convention adds a large amount of saturated color on the screen, which may mask information that is intended to be more salient (such as alarms).

Before committing to a design, it is important to fully understand the impact of specific design decisions. Standards and guidance such as NUREG 0700 (NRC 2002, NUREG-0700) have sections that address color scheme, but much of the information is open to interpretation. Further, recommendations such as maintaining population stereotypes have a sound basis, but are made without a full understanding of the specific context in which they will be applied. For example, there is very little basis for some population stereotypes such as the use of red and green to identify valve position aside from the fact that that is how it is typically represented in the commercial nuclear power industry. Beyond that, the red/green convention is inconsistent with other meanings of red and green including red identifying undesirable states such as danger or alarms. This is an example of where there is inherent conflict in the existing human factors guidance on use of color. The studies described in this report are designed to address this inconsistency and identify whether maintaining a design that is consistent with population stereotypes actually enhances performance, or whether there are alternative designs that yield performance that is better than the designs that maintain those stereotypes.

To facilitate investigation of the color scheme, the team developed four versions of the interface presented in Figure 3. There are two versions of background colors (a fully dull screen version and a version with colored flow paths) and two representations of component status (white/gray and red/green).

The overall design concept was evaluated in a dynamic scenario in which operators conducted a startup procedure. The specific design questions related to color scheme were investigated in more targeted experimental tasks referred to as micro tasks in this report. The rationale for this approach is that it is challenging to investigate specific design features in full-scale dynamic scenarios for several reasons. First, during full scale scenarios many factors contribute to performance, and the effect of specific variables such as a particular color choice is hard to separate from other factors such as scenario complexity or crew communication. Second, it is impractical to conduct full-scale studies that isolate specific variables such as specific color choices due to the amount of time that each scenario takes. Realistic scenarios can take anywhere from 20 minutes to hours to complete, and conducting enough trials to isolate the effect of specific variables may take days or weeks. Finally, design decisions such as the effect of color choice on performance can be initially tested in smaller scale studies, which are simpler and less expensive to conduct than full scale scenarios, and then validated in larger scale integrated system validation studies.

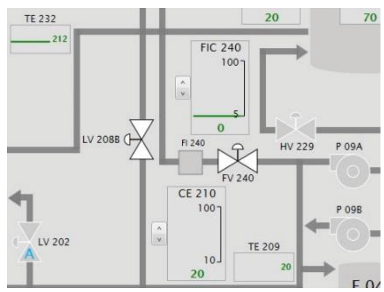
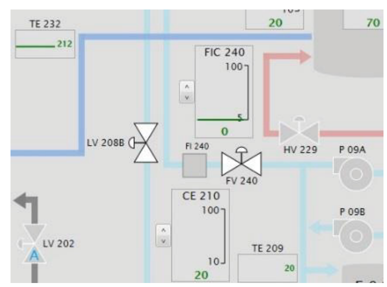
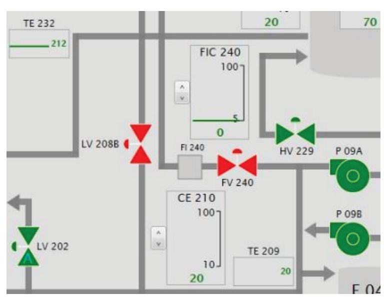
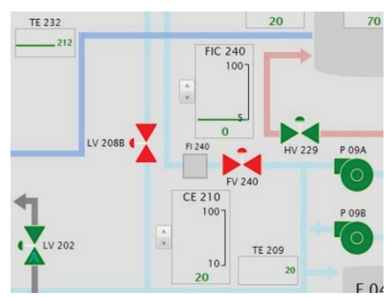
<table>
<tr><td></td><td></td><td colspan="2">Presentation of Flow Paths</td></tr>
<tr><td></td><td></td><td>A. Single color for all product streams/flow paths</td><td>B. Muted colors representing different product streams/flow paths</td></tr>
<tr><td rowspan="2">Presentation of Valves/Pumps</td><td>1. White gray and muted blue only</td><td>
</td><td>
</td></tr>
<tr><td>2. Red/green stereotype</td><td>
</td><td>
</td></tr>
</table>

Figure 3. Versions of interfaces developed to test color scheme.

# 2. Experimental Micro Task Study

One of the unanswered design questions for the radiological waste control room prototype was what color conventions should be used in the design. Specifically, the questions were whether to maintain the red/green stereotype for component status and whether adding background color to highlight different product streams would be beneficial to operators. Specifically, the micro tasks were designed to address the following research questions:

1. Do red/green component status indications help operators determine component state faster or more accurately than white/grey component status?
2. Do red/green valve and pump indications help operators more quickly determine how a system is aligned?
3. Do color coded product streams improve the operator's ability to identify consequences of system configuration more accurately or more quickly?
4. What are the interactions between the different color use strategies (i.e., do color coded product streams reduce the effect of red/green valves or the salience of alarms)?
5. Does the dull screen help the operator detect alarm states more quickly or accurately?

The following sections describe the method and results of the micro task study.

## 2.1 Method

### 2.1.1 Participants

A total of 12 auxiliary operators from the partner utility consented to participate in the Micro-Task experiments. Of these, two operators had to leave prior to completing participation, resulting in a total of ten operators with usable response time data (these operators were called away to conduct higher priority work, and would have remained in the study if they were available). Of these 10 operators, the eye tracker malfunctioned midway through the experiment for two additional operators resulting in a total of eight operators who yielded usable eye tracking data. Of these operators, 4 were fully trained to operate the LRS system, which represents the entire population of existing operators who would use the LRS system. The rest of the operators had some experience with the LRS system. Most operators had over 5 years of experience.

### 2.1.2 Experimental Design

There were two independent variables in this study: Component color and flow path color. Each had two levels and were manipulated within-participants. Figure 3 illustrates the different conditions. All participants saw all four versions of the interface represented in the 2 X 2 table in Figure 3, and the order each interface was presented in was counterbalanced using a balanced Latin square design.

The dependent variables were:

- *Accuracy*. Defined as the number of correct responses out of the total number of trials.
- *Response Time (RT)*. Defined as the total time elapsed from the trial onset to the time the participant responded.
- Eye tracking measures (for detailed descriptions, see Kovesdi, C., R. Hill, J. Oxstrand, Z. Spielman, K. Le Blanc, and T. Hansen):

- *Fixation count (FC).* Defined as the frequency of fixations (i.e., temporal and spatial pauses in eye movements where information processing occurs) during a trial. This s measure of visual search efficiency, and lower numbers indicate better efficiency.
- *Time to first fixation (TTFF).* Defined as the time difference between the trial onset to the time of the first fixation that landed on the target AOI (e.g., the information needed to answer the question). This s measure of visual search efficiency, and lower numbers indicate better efficiency.
- *LTLFTR*. Defined as the time difference between TTLF to the RT. This represents the amount of time between the last time the target AOI was looked at and when the response was submitted. This may represent the efficiency of the information processing that occurs between locating the information on the screen and comprehending what it means. Lower numbers indicate better performance.
- *Fixation duration (FD).* Defined as the average fixation duration (*ms*) for a given trial.

### 2.1.3 Apparatus

The micro tasks were presented as a series of questions using a stimulus presentation program developed by the Institute of Energy Technology. The program displays static images as visual stimuli and reads in the location and name of areas of interest (AOI) from extensible markup language (XML) data files. The AOIs were created as a square shape centered on the provided location. Using the location of the users gaze provided by the eye tracking technology and the given AOIs, the program recorded the frequency and duration that the AOIs were viewed. The program also recorded the order in which the operators fixated on the AOIs. Response times and accuracy based on key presses were recorded in the program. The AOIs were defined based on the components displayed on the evaporator overview (i.e., pumps, valve, controllers and trends) and the target AOI was defined as the component that was identified in the question for each trial.

### 2.1.4 Tasks

The main task was to respond to a questions based on the image presented on each trial. The image was a screen shot of the evaporator system, and the configuration of the evaporator system varied e.g., valve positions and pump status changed) to correspond to each question. The questions were developed by an expert with operations experience and knowledge of the evaporator system and a human factors scientist. The questions were arranged in blocks that were designed to address each of the research questions presented in section 2 as described below.

- Research Question 1: Do red/green component status indications help operators determine component state faster or more accurately than white/grey component status indications?

  This research question was addressed by a block of trials that contained questions that ask about the status of a single component. E.g., what is the position of [valve X]?”

  Hypothesis: Operators will respond faster (lower RTs) and more accurately (higher accuracy) when the components preserve the red/green stereotype. Eye tracking measures will also show a benefit to red/green component status (i.e., FC, FD, and TTFF will all be lower for red/green versus white grey components)

- Research Questions 2, 3, and 4: 2) Do red/green valve and pump indications help operators more quickly determine how a system is aligned? 3) Do color coded product streams improve operator's ability to identify consequences of system configuration more accurately or more quickly?, and 4) What are the interactions between the different color use strategies

  Questions 2, 3, and 4 were addressed with trials that asked about the consequence of changing a component status (e.g., opening a valve) on a parameter in the system (e.g., a level in a tank) an example question is "With the system in the current configuration what would be the result on level of [tank y] if we [open/close] [valve X]?"

  Hypothesis: There will be an interaction between component color and flow path color, with maximal performance observed for the HSI with dull components and colored flow paths. RT will be lower for that HSI and accuracy will be higher. Eye tracking measures will also show a benefit to dull components/colored flow paths compared to the other HSIs (i.e., FC, FD, and TTFF will all be lower for dull components/colored flow paths compared to the other HSIs). Finally LTLFTR will be lower for the dull components/colored flow paths compared to the other HSIs

- Research Question 5: Does the dull screen help the operator detect alarm states more quickly or accurately? This question was addressed with alarm trials that were embedded into each questions block. The participants were instructed to look for and respond to alarms regardless of the question asked.

  Hypothesis: Operators will respond faster (lower RTs) and more accurately (higher accuracy) with the dull screen. Eye tracking measures will also show a benefit dull screen (i.e., FC, FD, and TTFF will all be lower)

### 2.1.5 Experimental protocol

Each of the four HSI designs (presented in Figure 3 above) were tested in an experimental block. The order the HSI blocks were tested in was counterbalanced across participants using a balanced Latin squares design to control for order effects. Within each HSI block, there were a total of three question blocks (i.e., Block 1, Block 2, and Block 3); these question blocks contained multiple trials. For each HSI type, each participant completed the question blocks 1, 2, and 3 sequentially (i.e.., the order of question blocks was the same for each participant and for each HSI type, only the HSI presentation order varied). Individual questions within each block were randomly assigned each trial without replacement.

The questions in each of the blocks were progressively more complex. Block 1 contained an 'A' and 'B' sub-block. Block 2 also contained an 'A' and 'B' sub-block. Block 3 did not contain any sub-blocks. An embedded alarm task was presented in all blocks. The rationale for embedding the alarm detection experiment as a secondary task was to provide additional realism. That is, it was assumed that having alarms embedded as a secondary task would be a more valid assessment of the efficacy of the alarm's visual salience qualities versus merely instructing participants to find the alarm on the display due to the fact that operators must detect alarms in complex condition under some uncertainty in real-world operation tasks. Table 1 provides a summary of question blocks and examples.

Table 1. Summary of question blocks.

| Block | Sub-Block | Question Type | Number of Trials |
|---|---|---|---|
| 1 | A | "What is the position of [valve X]?" | 10; 3 alarms |
| | B | "What is the status of [pump X]?" | 4; 1 alarm |
| 2 | A | "What system is [component X] associated with?" | 7; 3 alarms |
| | B | "What would be the effect of changing the component status on the temperature in [tank y] by [opening/closing] [component X]?" | 5; 2 alarms |
| 3 | N/A | "With the system in the current configuration what would be the result on level of [tank y] if we [open/close] [valve X]?" | 7; 2 alarms |

The questions were designed to yield binary responses (e.g., closed or open, and on or off) and participants responded using either the 'z' or '/' key to answer each question. Participants responded to alarms by pressing 'space' rather than answering the question. Response times and accuracy were collected from these key presses. Eye tracking was also used to measure visual search behavior.

Each question set block contained a brief familiarization period in which the participant was instructed to press each key (e.g., 'z, /, space') to map to the corresponding response to the appropriate key (e.g., 'open, close, alarm'). Each question set block contained a similar workflow where a question was presented in the top center of the screen until the participant read and acknowledged the question by pressing 'enter.' Next, a brief mask was presented with a crosshair in the center of the screen for 1000 milliseconds (ms). The purpose of the fixation screen was to ensure the participants' initial gaze was in a consistent location prior to starting each trial. Finally, a stimulus image was presented along with the question in the top center of the screen. Participants were instructed to answer as quickly and accurately as possible, and to prioritize finding alarms over answering the question. Figure 4 illustrates this workflow

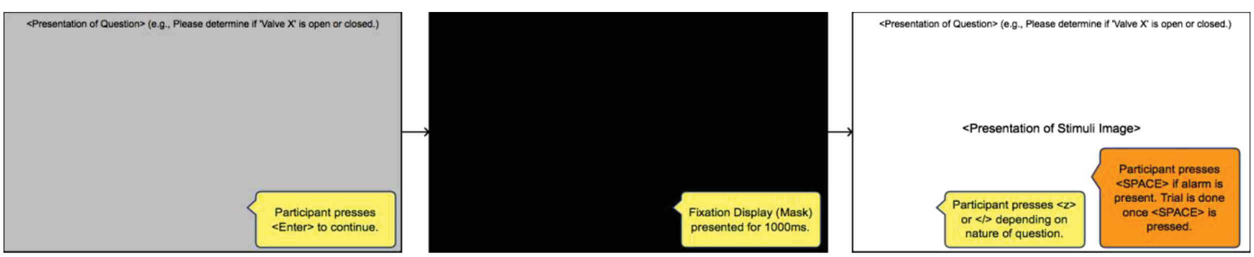



Figure 4.Overall Trial Design for the Micro Task Studies.

## 2.2 Results and Discussion

### 2.2.1 Analysis

A multi-level model (MLM) was created for response times and eye tracking measures. One motivation for using the MLM was its ability to handle dependent data (i.e., by participant). Likewise, MLM is more robust in handling instances of homogeneity of regression slopes, assumptions of independence, and missing data, which would otherwise violate assumptions common to traditional general linear models like analysis of variance (Field et al., 2012). Each MLM model specified independent variables *HSI Order*, *Component Color*, *Flow Path Color* and *Component*Flow Path Color* as random effects, being nested within participant. Likewise, each MLM model sequentially introduced independent variables to systematically test the statistical contribution of each independent variable. Post hoc tests were run on each independent variable that yielded statistical significance using Tukey Honest Significant Difference post hoc tests. All MLM models were run in R. The results reported are all significant at the alpha = .05 level. An example MLM model in R is provided in Figure 5.

```r
library(nlme)
baseline<-lme(Avg_Fixation_Duration~1,random = ~1|Participant/HSI_Order/Component/FlowStream,data=FD_B1,method="ML")
model_HSI_Order<-update(baseline,.~.+HSI_Order)
model_Component<-update(model_HSI_Order,.~.+Component)
model_FlowStream<-update(model_Component,.~.+FlowStream)
model_Interaction<-update(model_FlowStream,.~.+Component:FlowStream)

anova(baseline,model_HSI_Order,model_Component,model_FlowStream,model_Interaction)
```

Figure 5. MLM performed in the R Statistical Computing Environment.

For response time and eye tracking results, error trials were excluded from the analysis.

The overall counts of correct responses to the questions alarm trials were analyzed using Fisher's exact test.

### 2.2.2 Block 1: Simple Component Status

Block 1 was designed to answer the research question: Do red/green component status indications help operators determine component state faster or more accurately than white/grey component status? Block one contained simple questions such as, what is the position of [valve X]?" The researchers hypothesized that operators would respond faster (lower RTs) and more accurately (higher accuracy) when the components preserve the red/green stereotype. Eye tracking measures will also show a benefit to red/green component status (i.e., FC and TTFF will all be lower for red/green versus white grey components).

The MLM analysis of response time and eye tracking data revealed a significant difference between component color for response time and fixation count, with colored components yielding better performance than dull components (see Figure 6). There were no differences in TTFF and there were no effects of flow path color on response times or eye tracking measures.

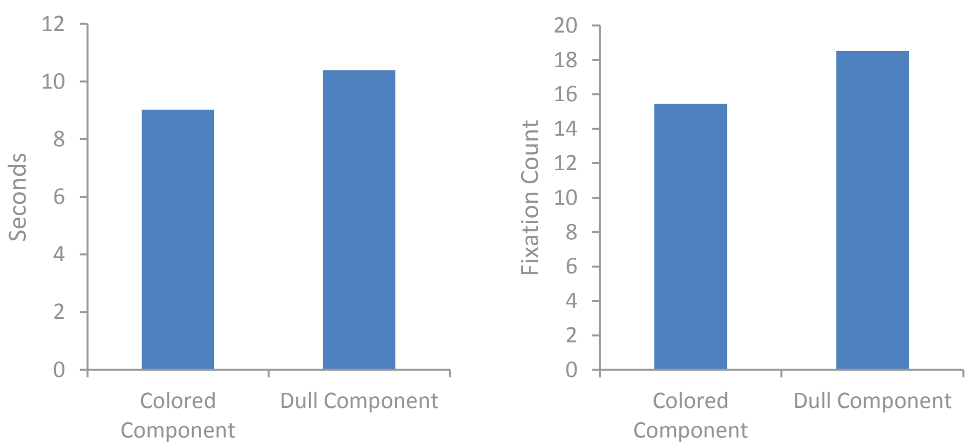



Figure 6. Mean Response Time (RT) and Fixation count (FC) for Block 1

There was also a significant difference based on flow path color for fixation duration Figure 7 shows that fixation duration was greater when there were colored flow paths compared with grey flow paths

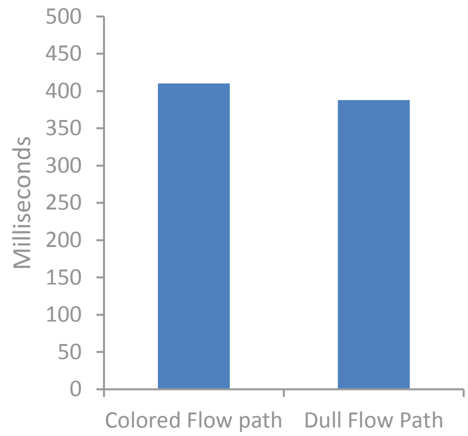



Figure 7. Mean fixation duration based on flow path color.

Fisher’s exact test on the accuracy data revealed a significant effect of component color, with fewer errors occurring with dull components than with colored components. Figure 8 illustrates the difference.

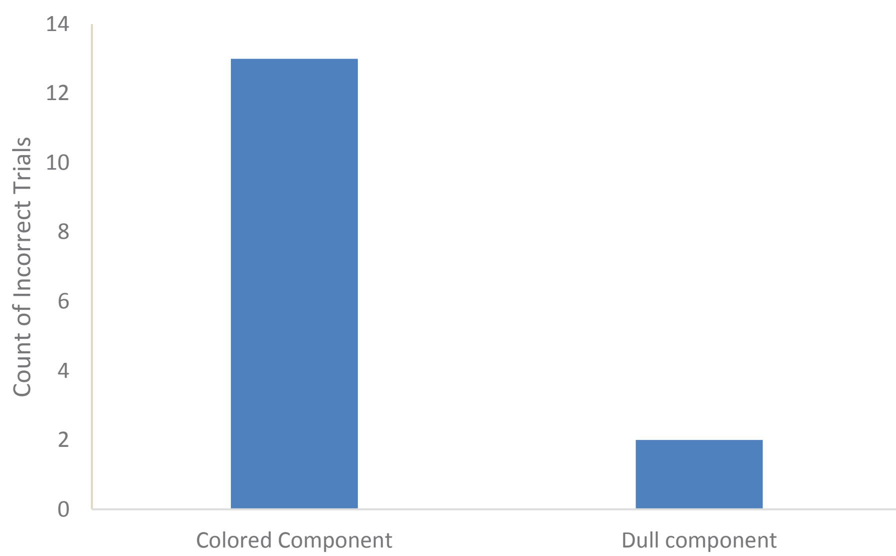



Figure 8. Total count of incorrect trials for block one by component color.

Response time and eye tracking data showed an advantage of having red/green components, however the faster response times came at the cost of lower accuracy. The results of block 1 indicate that the performance advantage of using red/green components is minimal for questions that simply ask about the component status, and the increase in speed and decrease in visual searching required may not be worth the reduction in accuracy. Further, in most cases, an operator's task is not simply to determine the status of a single component, but also to determine how a system is aligned based on multiple component statuses and to determine the impact of changing the position of a valve or starting/stopping a pump. Blocks 2 and 3 were designed to address these more complex questions. With the exception of the accuracy results, which favored the dull components, these findings are consistent with the hypothesis.

### 2.2.3 Blocks 2 and 3: System Alignment and Consequence of Actions

Blocks 2 and 3 were designed to address the following research questions:

- Do red/green valve and pump indications help operators more quickly determine how a system is aligned?
- Do color coded product streams improve operator's ability to identify consequences of system configuration more accurately or more quickly?
- What are the interactions between the different color use strategies?

The researchers hypothesized that there would be an interaction between component color and flow path color, with maximal performance observed for the HSI with dull components and colored flow paths. They hypothesized that RT will be lower for dull components/colored flow paths and accuracy will be higher. Eye tracking measures will also show a benefit to dull components/colored flow paths compared to the other HSIs (i.e., FC, FD, and TTFF will all be lower for dull components/colored flow paths compared to the other HSIs). Finally LTLFTR will be lower for the dull components/colored flow paths compared to the other HSIs.

Although the questions in blocks 2 and 3 were designed to address the same research hypothesis, the questions in block three were more challenging, so the question blocks were analyzed separately. There were no statistically significant effects of component color or flow path color for Block 2A, which had questions that asked what system a particular component was associated with.

For Block 2B there was a significant interaction between flow path color and component color as measured by both response time and fixation count, with the greatest performance occurring for colored flow paths and dull components for both FC and RT (see Figure 9).

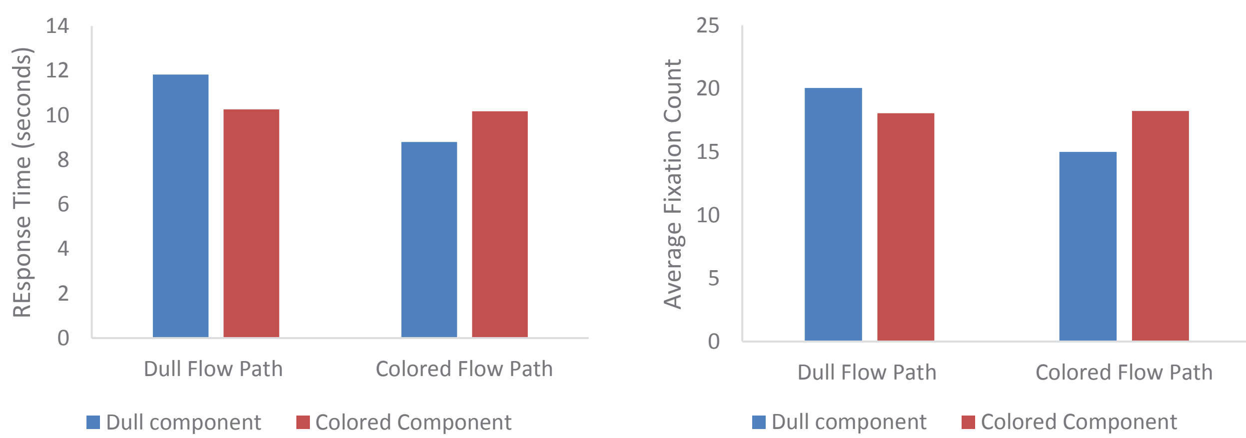



Figure 9. Average response time and fixation count for Block 2B.

These results suggest that the use of red/green component coloring may reduce any benefit of flow path color. When component colors were dull, flow path coloring showed improved visual search performance and efficiency compared to all gray colored flow paths. As shown in Table 1 the questions provided in Block 2B required integration of content from the HSI displays in order to answer the question. It was particularly interesting that the results here suggest that when component colors are muted (gray/white), then the used of colored flow paths improve visual search performance.

There also was a statistically significant interaction effect (i.e., Component Color*Flow Path Color) for Fixation Duration indicating that the HSI with Colored Components and Dull Flow paths had a lower fixation duration compared to all other HSIs (See Figure 10).

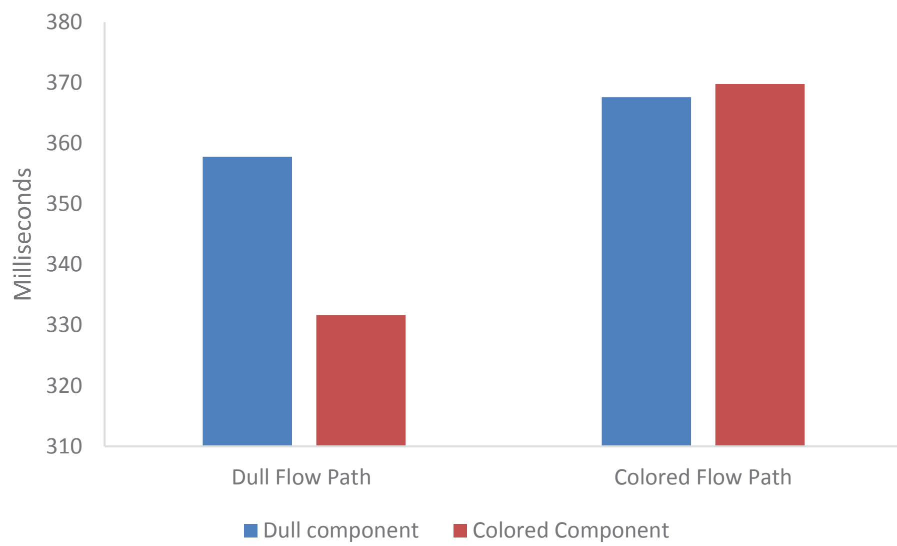



Figure 10. Fixation Duration for Block 2B.

This result suggests that when colored components were provided with dull flow paths, there was less cognitive effort observed in visual search. This result may be indicative of having a familiar color scheme for operators (e.g., red/green) while reducing the amount of information available on the display (i.e., by muting flow paths). It should be noted that while fixation duration was least in this condition, this design did not yield any better visual search performance, indicating that although the operators spent less time focusing on elements of the screen, they did not find the relevant information any more quickly.

For Block 3, the MLM analysis revealed a significant interaction between flow path color and component color of Response Time with the effect of flow path color depending on the component color. Figure 11 shows that colored flow paths with dull components yielded the best response time performance.

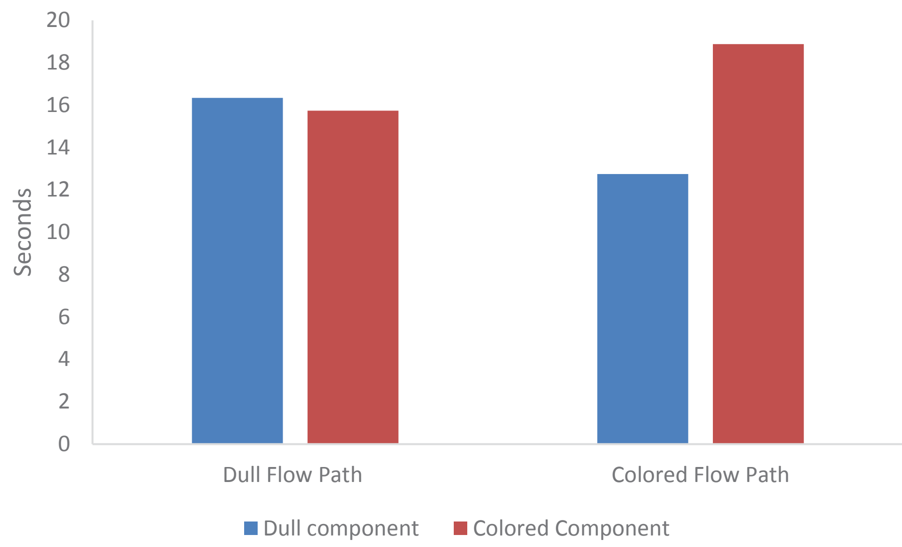



Figure 11. Average response time for block three.

The MLM also revealed a significant interaction for fixation count, with colored flow paths providing a benefit with dull components, but not colored components as shown in Figure 12.

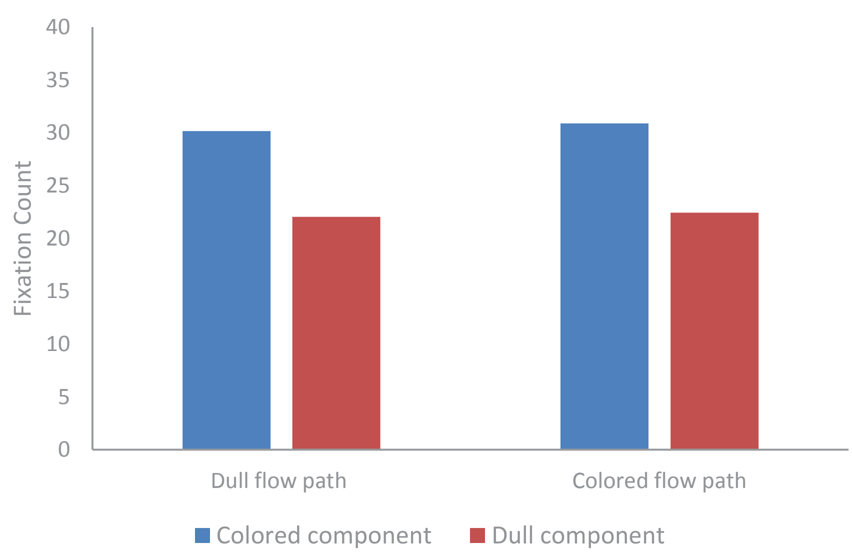



Figure 12.Average fixation count for block 3.

There was a statistically significant main effect found for Component Color with LTLFTR, indicating that lower latencies between the final target fixation to response time was found with HSI designs with dull component colors. This may indicate that with dull components the amount of time that they had to think about or comprehend the information to determine the answer was lower, indicating more efficient information processing with dull versus colored components (see Figure 13 ).

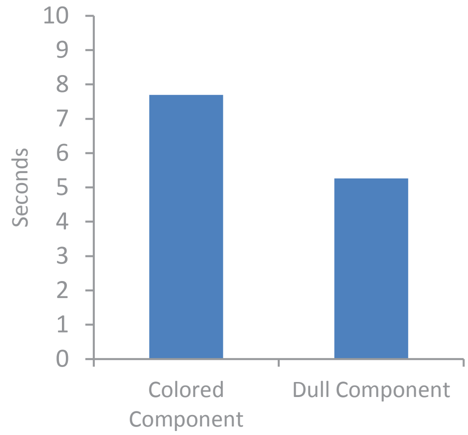



Figure 13. Average LTLFTR for block 3.

#### *2.2.3.1 Other Observations*

It should also be noted that learning effects were found for Block 1 and Block 3, where visual search performance was improved through order. It should be emphasized that HSI block order was included as a random variable in the MLM in order to statistically control for this effect on component and flow path effects.

### 2.2.4 Alarms

The final research question was addressed with the alarm trials. Specifically the question was: Does the dull screen help the operator detect alarm states more quickly or accurately? The researchers hypothesized that operators would respond faster (lower RTs) and more accurately (higher accuracy) with the dull screen.

Analysis of response time data showed no significant effects of HSI on Alarm trials. However a Fisher exact test on the accuracy counts revealed a significant effect of component color on detection of alarms, with dull components showing greater detection of alarms than colored components (see Figure 14).

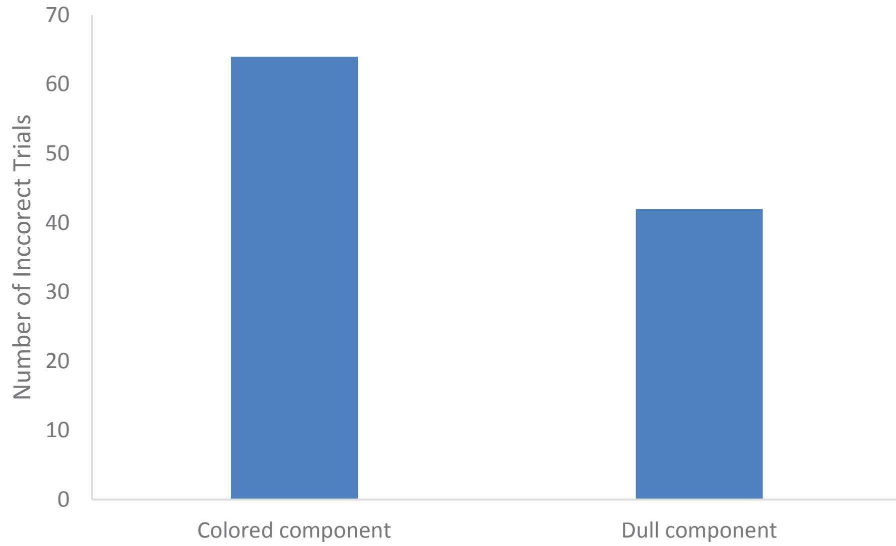



Figure 14. Alarm trial accuracy by component color.

The Fisher Exact score revealed a marginally significant difference between the four versions of the HSI ($p = 0.057$), showing that the HSI with dull components and dull flow paths had the fewest missed alarms (Figure 15).

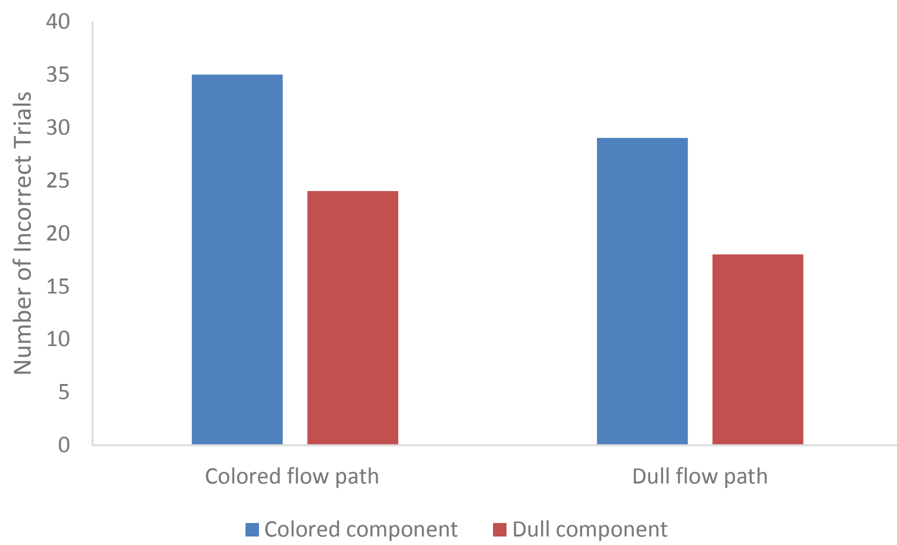


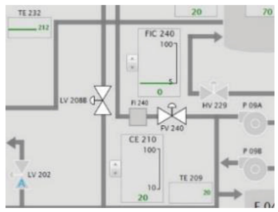
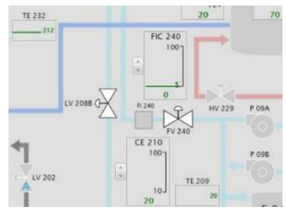
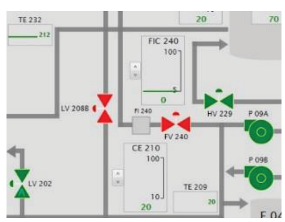
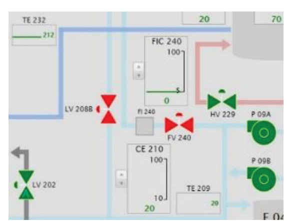

Figure 15. Alarm trial accuracy by component color and flow path color.

The alarm results indicate that colored components may influence the number of alarms that are missed, but color does not seem to affect how quickly alarms are detected. The results are consistent with the hypothesis in that dull screen facilitate detection of alarms. The red color used to indicate component status was the same as the red used to indicate alarms, therefore it makes sense that on displays with colored components, the alarms state was not as easily detected.

## 2.3 Summary of Micro Task Results

Table 2 summarizes the findings of the micro task study by HSI type.

Table 2. Summary of micro task results

| | | Flow Path color | | |
|---|---|---|---|---|
| | | Single color for all product streams/flow paths | Muted colors representing different product streams/flow paths | Component color Results |
| **Component color** | White gray and muted blue only | <br>✓ Best Alarm detection performance | <br>✓ Fastest response time for identifying system alignment and consequences<br>✓ Most efficient visual search time for identifying system alignment and consequences | ✓ Most efficient "comprehension time" |
| | Red/green stereotype |  |  | ✓ Fastest Response times and better search efficiency for identifying component state<br>× Lowest Accuracy when identifying component state |

## 2.4 Conclusions and Recommendations

The micro task study revealed that colored components (i.e., red/green) yield slightly faster response times when the task is to identify just the component status, but there is no benefit to colored components when the task is more complex (e.g., identifying the consequence of changing a component status with a given system alignment). Further, although response times were faster for identifying component status with the red/green color scheme, accuracy was lower, so the greater efficiency came with a tradeoff. Overall, the eye tracking, accuracy and response time results yielded the most positive conclusions for the HSI with dull components and colored flow paths as performance on visual search, and response times were best for the questions when using that HSI.

The results suggest that there is little basis for maintaining the red/green stereotypes, and that by using an alternative design, operator performance might actually improve. The results also suggest that if a design uses saturated color to highlight alarm states or provide alerts, then using that same saturated color for other design elements will reduce the efficacy of the alarm color. In general, the results support using muted colors for static background information and refraining from using saturated colors for everything but alarm information. The results don't support using a purely dull screen approach as the colored flow paths provided tangible performance benefits when determining how a system is aligned and what the consequences of changing a component status would be.

The results also suggest that the dull screen concept (i.e., no saturated color for static information) may produce the best alarm detection performance, so for situations for which abnormal states need to be highlighted and most information in steady state operation doesn't change, the dull screen approach may be beneficial.

Generally, color is effective at conveying important information, but as the amount of saturated color increases, the value of the information decreases and search efficiency and response time suffer. The researchers recommend using color sparingly, choosing muted colors for static background elements and reserving saturated color for highlighting abnormal states and other important dynamic information.

# 3. REFERENCES

Boring, R. L., V. Agarwal, J. C. Joe, and J. J. Persensky, *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 National Laboratory, 2012.

Boring, R. L., J. C. Joe, T. A. Ulrich, and R. T. Lew, "Early-stage design and evaluation for nuclear power plant control room upgrades," *Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting*, October 27–31, 2014, pp. 1909–1913.

Field, A., Miles, J., & Field, Z. (2012). Discovering statistics using R. Sage publications.

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

Kovesdi, C., R. Hill, J. Oxstrand, Z. Spielman, K. Le Blanc, and T. Hansen. Prototype Design, Analysis, and Results for a Liquid Radiological Waste Control Room. INL/EXT-17-43226 Revision 0, Idaho National Laboratory, 2017.

Le Blanc, K., R. L. Boring, J. C. Joe, B. Hallbert, and K. Thomas. "A Research Framework for Demonstrating Benefits of Advanced Control Room Technologies." INL/EXT--14-33901, Revision 1, Idaho National Laboratory, 2014.

NRC 2002, NUREG-0700, "Human-System Interface Design Review Guidelines," Rev. 2, U.S. Nuclear Regulatory Commission, 2002.

Ulrich, T., R. Boring, and R. Lew, *Human Factors Engineering Design Phase Report for Control Room Modernization*, INL/EXT-14-33221, Idaho National Laboratory, 2014.