RIL 2022-11

Human Performance Test Facility Volume 5

Touchscreen Ergonomic Considerations

Date Published: August 2025

Prepared by:
K. Dickerson
N. Hughes Green

U.S. Nuclear Regulatory Commission
Office of Nuclear Regulatory Research
Division of Risk Analysis
Human Factors and Reliability Branch

Kelly Dickerson, NRC Project Manager

Research Information Letter
Office of Nuclear Regulatory Research

Disclaimer

Legally binding regulatory requirements are stated only in laws, NRC regulations, licenses, including technical specifications, or orders; not in Research Information Letters (RILs). A RIL is not regulatory guidance, although NRC's regulatory offices may consider the information in a RIL to determine whether any regulatory actions are warranted.

PREFACE

Much of the basis for current U.S. Nuclear Regulatory Commission (NRC) Human Factors Engineering guidance comes from research conducted in other domains (e.g., aviation, defense), qualitative data from operational experience in nuclear power plants, and a limited amount from empirical studies in a nuclear environment. The Commission, in Staff Requirements Memoranda SECY-08-0195, approved the staff's recommendation and directed the staff to consider using generic simulator platforms for addressing human performance issues, as simulators provide a tool to gather more empirical nuclear specific human performance data. These data are intended to enhance the current information gathering process, thus providing stronger technical bases and guidance to support regulatory decision making. To that end, in the spring of 2012, the NRC sponsored a project to procure a low-cost simulator to empirically measure and study human performance aspects of control room operations to address the human performance concerns related to current as well as new and advanced control room designs and operations. Using this simulator, the Human Factors and Reliability Branch in the Office of Nuclear Regulatory Research Division of Risk Analysis began a program of research known as the NRC Human Performance Test Facility to collect empirical human performance data with the purpose of measuring and ultimately better understanding the various cognitive and physical elements that support safe control room operation. Additionally, the baseline methodology documented in these volumes will enable Human Reliability Analysis (HRA) data research that will address key gaps in available data for topics such as dependency and errors of commission, improving the state of the art of HRA and thus dual human factors and HRA data missions.

This Research Information Letter (RIL), 2022-11, Volume 5, presents a reanalysis of a subset of the data reported in Volumes 1 and 2 of the same RIL series. The focus on this target re-analysis is on the potential for ergonomic issues associated with the use of touchscreens. This is because there is increased interest in using touchscreens (i.e., glass top simulators) for training, licensing exams, and validation and verification activities. Some of this is due to the practicalities of plant restarts, as a restarting facility may not have fully established their plant referenced simulator at the time the first wave of licensing exams must take place. Additionally, advanced reactor control rooms may use touchscreens for non-safety related systems. The preliminary findings presented in this limited scope re-analysis are the first step in documenting the potential limitations, use cases, and advantages of glass top simulators. These results will support licensing exams and review activities for both large light water reactors, small modular reactors, and advanced reactors.

ABSTRACT

There has been significant interest in using touchscreens (i.e., glass top simulators) for operator examinations, training, and potentially validation and verification activities. Glass top simulators differ from the plant and are, therefore, not considered a plant-referenced simulator that must meet the Commission requirements outlined in 10 CFR 55.46. Glass top simulators can represent the plant in high fidelity, with one key difference – the interaction modality. The research presented in this report is a re-analysis of a subset of data from RIL 2022-11, Volumes 1 and 2 and focuses on the workload and performance differences between a desktop workstation, glass top, and hard panel simulator. The research includes survey-based, physiological, and performance-based measures of workload in student (novice) and operator (expert) populations on a series of representative control room tasks. As in previous Human Performance Test Facility studies, there was a difference between self-rated workload and physiological measures, with participants (both experts and novices) underestimating their workload relative to the physiological indices. Additionally, both operators and students had to make repeated attempts to execute required actions. There was also evidence of ergonomic factors, specifically the position of controls on the touchscreen as potential factors driving the need for the repeated attempts to complete the same action. In general, it was observed that participants had to make more attempts to execute actions in the extreme upper and lower screen positions than for more centrally located controls. While this points to a potential limitation of touchscreens, additional research is needed as these analyses were post-hoc and not the central focus of the studies where the data were sourced. The findings are discussed in the context of operator training and examination; however, they are also applicable to advanced reactor control room operations. The plan for future research is also discussed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

Human Factors Reviews Overview

The staff of the U.S. Nuclear Regulatory Commission (NRC) are responsible for reviewing and determining the acceptability of new reactor designs or modifications to operating plants. These reviews ensure that new designs or design modifications support safe plant operations. To perform these reviews, the staff use Chapter 18 of the Standard Review plan (NUREG-0800) and the guidance documents referenced therein, specifically the NRC Human-System Interface Design Review Guidelines (NUREG-0700, Revision 3[1]) and the Human Factors Engineering Program Review Model (NUREG-0711, Revision 3). NUREG-0700 addresses the physical and functional characteristics of human-system interfaces (HSIs) and provides guidelines for the performance of HSI or HSI component evaluations. The research discussed in this report is most directly relevant to issues described in NUREG-0700.

NUREG-0700 is a comprehensive guide that describes the physical and functional hardware and software characteristics of control room HSIs. The document also provides high level HSI design review criteria covering everything from automation displays, communication systems, and even chairs, footrests, and the physical dimensions of control labels. Chapter 11 of NUREG-0700 focuses on issues related to workstation design. Section 11.3, “Workstations Containing Primarily Computer-Based HSIs” covers the design characteristics of computer-based HSIs. This includes both conventional (i.e., keyboard and mouse) and touchscreen workstations. Section 11.3.2.3.6 describes circumstances where a touchscreen device is not recommended. This includes circumstances where the operator would have to hold their arm up to the screen for a prolonged interval or enter a large amount of data. Section 11.3.2.3.6 also indicates that frequent switching between input devices could present a challenge for operators. Section 11.3.2.3.6, describes several use cases where touchscreens may be desirable, for example, occasionally selecting a specific graphical object, interacting with elements on a layout or piping diagram, or opening/closing valves represented on digital controls. Section 11.3.2.3.6 provides a summary table of both advantages and disadvantages associated with implementation of touchscreens in a control room environment (see Table 1). While these recommendations are comprehensive, they miss a few key considerations revealed by more recent research with large format touchscreens, specifically spatial position effects, gesture type challenges, and the role of operator body size in driving performance during large format touchscreen use.

Table 1. Advantages and Disadvantages of Touchscreens

Advantages	Disadvantages
No separate input devices	Slower alphanumeric data entry

[1] At the time of writing this report, the next update to NUREG-0700, Revision 4 was being finalized. Substantive updates relevant to this work include Chapter 7, “Soft Control Systems” and a new Chapter 15, “Integration of HSI Resources”.

Programmable interface	Arm fatigue
Fast access	Finger may obstruct view
Direct manipulation of targets	Fingerprints or other debris may obscure screen
Input/output in same location	Large button required for finger use
Intuitive	Pointing is not very accurate
Natural pointing action	User must be within reach of screen
Generally, no additional desk space required[1]	No tactile feedback provided[3]
Generally, no training required[2]	Unable to rest finger on target without actuation[3]
	Gloved operation may be incompatible with some touch technology
	Controls must be deactivated for cleaning

Note: 1. If incorporation is part of an existing primary display. 2. Application dependent. 3. If a tactile membrane is not incorporated. Table source: NUREG-0700, Revision 3 Table 11.7.

As HSI technologies and plant modernization advance, touchscreens may play an increasingly important role in future nuclear power plant (NPP) simulators and control rooms. However, as Table 1 demonstrates, there is also potential for human performance and ergonomic issues.

Is There a Use Case for Touchscreens?

It is not yet clear if there will be wide use of touchscreens or glass top for operations in control rooms. However, there has been increased interest in the industry in incorporating touchscreens into their licensing and training processes. For example, if a plant is newly built and their plant referenced simulator is not yet available, there may be a need to use a glass top simulator. Glass top simulators are capable of representing the full data content and data dynamic fidelity (see Table 2) of a plant's control room but use a distinct interaction design compared to the actual control room. This non-plant referenced simulator would be permissible from a regulatory perspective if it is suitable to conduct an operating test as outlined in 10 CFR 55.45(a). Applicants, when opting to use a non-plant referenced simulator, generally want the simulator to look and function as close to the actual control room as possible. The pursuit of maximum fidelity leads to a situation where large format touchscreens, such as those used in glass top bays, require specific gestures and manual force to operate effectively. These actions are not aligned with the typical capabilities and usability of touchscreens. This can lead to a mismatch between fidelity and usability and potentially lead to challenges for license examiners in determining if an error made during an exam is due to a limitation in executing a complex manual gesture or a lack of plant system knowledge. Additionally, if future licensed operators, particularly a plant's initial group receives all their simulator-based training using the touchscreen system, they may lack critical knowledge about the way to engage the controls and how much force to apply, particularly hand turned control mechanisms, switches requiring very small incremental changes (e.g., turning a dial to roll turbines), and large valves with long stroke times. This inherent functional fidelity limitation may lead to a need for supplemental training for controls that require complex gestures or a specific level of force to correctly operate.

Table 2. NUREG-0700, Revision 3 Fidelity Definitions.

Fidelity	Definition
Completeness	Testbed should represent the complete integrated system, including HSIs and procedures not specifically required in the test scenarios.
Physical Fidelity	HSIs and procedures should be represented with high physical fidelity to the reference design, including presentation of alarms, displays, controls, job aids, procedures, communications equipment, interface management tools, layout, and spatial relationships.
Functional Fidelity	HSIs and procedure functionality should be represented with high fidelity. All HSI functions should be available.
Environmental Fidelity	High environmental fidelity relative to the reference design, including expected levels of lighting, noise, temperature, and humidity.
Data Content Fidelity	Data content should be represented with high physical fidelity to the reference design. The presentation of information and controls should rest on an underlying model accurately mirroring the reference plant. The model should provide input to the HSI such that the information accurately matches that which is presented during operations.
Data Dynamics Fidelity	Data dynamics should be represented with high fidelity to the reference design. The process model should be able to provide input to the HSI so that information flow and control responses occur accurately and within the correct response time, e.g., information should be sent to personnel with the same delays as occur in the plant.

Ergonomic Considerations for Large Format and Bay Style Touchscreens

Functional fidelity is not the only usability limitation for glass top simulators. Interactions with touchscreens, particularly those in the vertical “bay” configuration have several potential usability challenges. One of the most visible is that large touchscreens require use of unsupported mid-air gestures. Unsupported mid-air gestures are defined as those where the arm has no external support as the operator engages the HSI. These gestures are notorious for inducing fatigue, quickly (Hansberger, 2017). Unsupported mid-air gestures are not the only ergonomic limitation. Davis, Hammer, Kotowski, and Bhattacharya (2014) compared different bodily postural positions for touchscreen and keyboard and mouse data entry conditions. In the touchscreen condition they found participants had the most discomfort in standing positions with low vertical and horizontal screens (Figure 1, panels G and K). While the nuclear control room context of use is different than the Davis et al. study, the example postural demonstrations suggest that controls in the lower portion of a large format touchscreen could create situations where an operator experiences discomfort or has reduced gestural control, potentially producing errors.

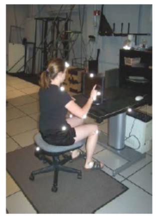
G	**H**	**I**	**J**	**K**	**L**
					
SVH	**SVL**	**CVL**	**SHH**	**SHL**	**CHL**
Standing, Vertical Screen, High Surface	**Standing, Vertical Screen, Low Surface**	**Chair, Vertical Screen, Low Surface**	**Standing, Horizontal Screen, High Surface**	**Standing, Horizontal Screen, Low Surface**	**Chair, Horizontal Screen, Low Surface**

Figure 1. Touchscreen Configurations Tested in Davis et al., (2024)

In the HSI design space, Breuninger (2020) found that the gesture-based interaction concept and the special surface characteristics of touchscreens create user challenges related to input feedback, size of interactive elements, compatibility, effects of virtual physics, and interference.

Table 3. Touchscreen HSI Considerations and Challenges.

Touchscreen HSI Considerations	**Description**[2]
Occlusion	Finger or portion of the hand can block required controls.
Feedback	Touchscreens lack mechanical buttons and do not provide feedback about the success of an action. Haptic feedback is not present by default and must be engineered into the interaction design. The lack of haptic feedback from the solid surface is usually addressed by adding vibration actuators, however, users generally rely on visual feedback.
Precision	For experienced users, action accuracy with conventional “touchscreen gestures”, such as point, pinch to zoom, and swipe to scroll is comparable to a mouse and trackball. However, the “fat finger problem” is a consideration. Fingers activate a larger surface area than a mouse pointer and will be less precise. Use of stylus is an effective mitigation for precision issues (and occlusion).
Environmental vibrations	High vibration areas are not suitable for touchscreen data entry as motion (environmental or self-generated) reduces accuracy and increases physical fatigue as the user engages more muscles to stabilize the arm and hand during use.
Virtual keyboards	Virtual keys should be 19x19 mm, to avoid mis-key errors. NUREG-0700 suggests that virtual keyboards should not be used when a large amount of data entry is required.
Posture/ergonomics	Except for horizontal touchscreens, users must hold up the hand and arm for some duration, during input (i.e., unsupported mid-air gesture). There is more opportunity for shoulder, arm, and hand strain compared to mouse pointer input devices.

[2] Table information sourced from Breuninger (2020) and Orphanides and Nam (2017)

Complexity	Roughly equivalent for selection/activation between point and click, however, touchscreens usually involve many different types of gestures.
Layered menus	Cascading menus and tools that require multiple parameter selection across multiple tabs, toolbars, or menus are more difficult to navigate and require more physical effort on a touchscreen compared to a conventional workstation.
Force estimation	Users tend to press harder than necessary to activate touchscreens. Users that move (e.g., walk, bend) make more input errors and these errors are not mitigated by familiarity with gesture-based input.
Parallax errors	A subtype of visual targeting error where, because of the perceived angle of approach, the user misses the target, touching too low (see figure 2).

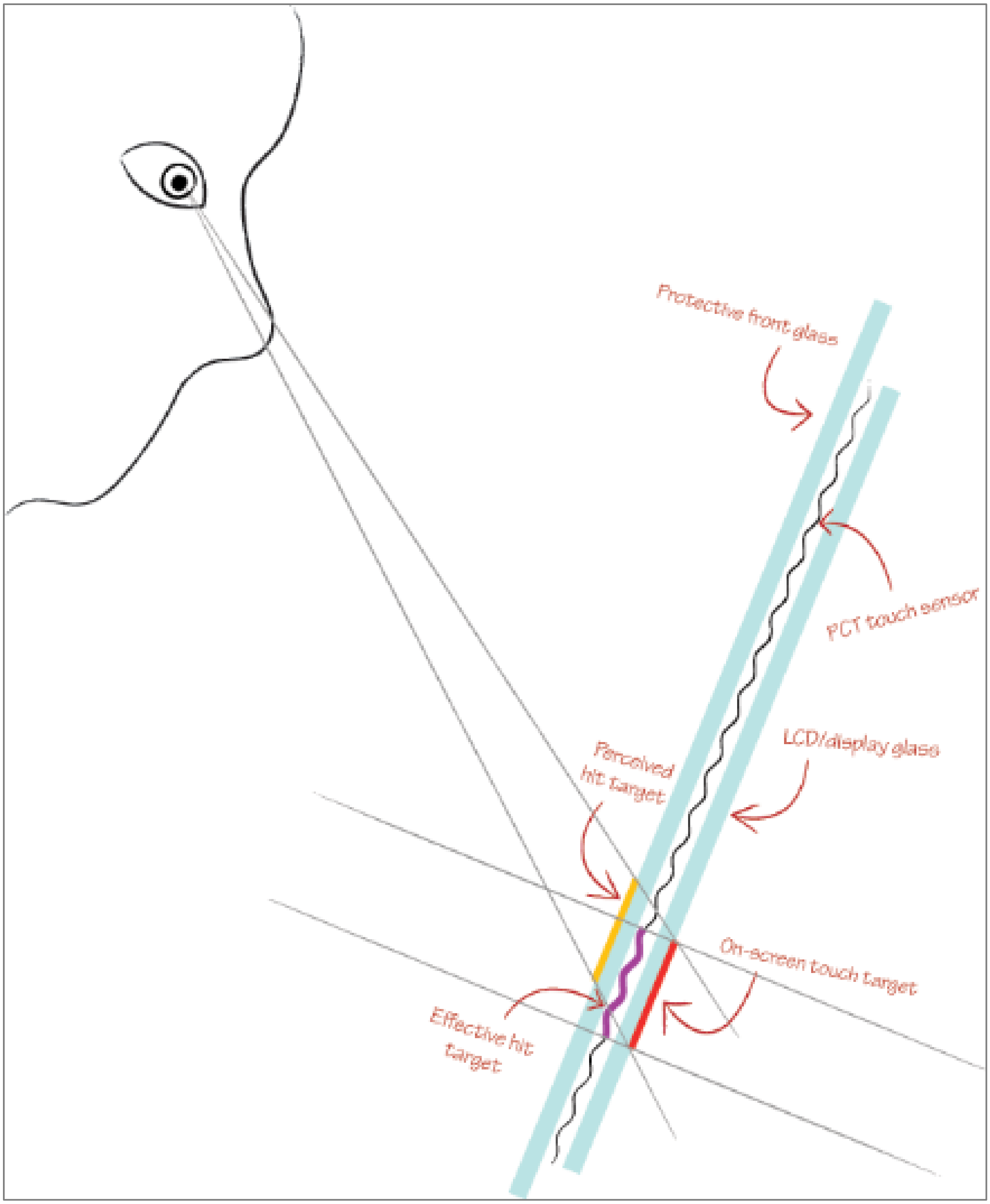


Figure 2. Example of a Parallax Error – User is targeting A but selects B because the touchpoint is actually lower than perceived because of the angle of approach3.

[3] Image source: https://ux.stackexchange.com/questions/22719/is-there-any-difference-in-element-minimal-sizes-on-10-and-24-inch-touch-screens

Re-Analysis for the Present Study

Prior Human Performance Test Facility (HPTF) studies have used a variety of interaction modalities (conventional keyboard and mouse workstation, touchscreen, hard panel) to support their HSIs. While the focus of those previous studies was not on differences between the various interaction designs, there were differences between modality, which, with the increased interest in using touchscreens from the industry, presents an opportunity to conduct targeted re-analysis of the previous studies. The focus of this re-analysis is to gain preliminary insights into the potential performance, usability, and ergonomic challenges introduced by touchscreens used in a glass top simulator.

To avoid duplicating reporting details published elsewhere in this RIL series, this report will only describe the methods related directly to the re-analysis. For a complete description of the history and motivation for the program, methods, and techniques used in the previous work, refer to RIL 2022-11, Volumes 1 and 2 (the reader will find sections 1.2, 2.2, and 2.4 particularly useful). Three features of the methodology are necessary for understanding the re-analysis, these are: (1) the use of both novice participants (i.e., university students) and operators, (2) the definition of task components and HSI modifications, and (3) the multivariate workload assessment, which included both subjective (i.e., surveys) and objective measures (performance, physiology).

Research Questions

This re-analysis is suited to address three questions about interaction modality.

1. Is there evidence of differences in workload between a keyboard and mouse and touchscreen input device?

2. Are there performance differences between a keyboard and mouse and touchscreen input device?

3. Is there evidence of differences based on the spatial position of controls?

CHAPTER 2 METHODOLOGY

Data Sourcing

Data were sourced from three previous studies (RIL 2022-11 Vol 1 studies 1 and 2; RIL 2022-11 study 2). The designs of the studies are summarized in Table 4. More detailed information is provided in the following sections. Due to a computer error during data collection with Reactor Operator 2 (RO2) in the student sample, only RO1 is analyzed for students.

Table 4. Data Source Summary

Study	RO1	RO2	Sample size	Simulator	Interface	Section in this Document
RIL 2022-11 Volume 1, study 1	Student Novice	Student[4] Experienced	69	GSE GPWR	Touchscreen Digital Desktop Workstation	2.2.1
RIL 2022-11 Volume 2, study 1	Former Operator	Former Operator	18	GSE GPWR Trojan	Touchscreen Digital Hard Panel Control Room	2.2.2

Procedure Selection, Task Classification and Analysis

All HPTF experiments use a modified generic Emergency Operating Procedure (EOP-EPP-001 GSE Power Systems, 2011). The procedure requires participants to perform predetermined tasks to respond to a loss of all alternating current power to the plant's safety buses. The EOP can be decomposed into a series of discrete tasks, labeled checking, detection, and response implementation. These tasks can be readily trained within the novice population as they are either rule or skill based and do not require detailed knowledge of the plant or a background in nuclear thermohydraulic. The selected tasks are also representative of those performed by ROs and directed by Senior Reactor Operators (SROs) (O'Hara et al., 2008; O'Hara & Higgins, 2010; Reinerman-Jones et al., 2013).

Checking requires a one-time inspection of an instrument or control to verify that it was in the appropriate state. *Response implementation* requires a motor response (mouse usage, finger touch, hand gesture) to change the state of the NPP by locating a control and subsequently manipulating it in the required direction. *Detection* requires continuous monitoring of a control parameter to identify a change in the state of the plant.

Tasks were composed of individual steps and each experimental scenario had twelve steps. The steps were grouped by task type (4 checking steps, 4 detection steps, and 4 response implementation steps). The task type order was partially counterbalanced to create scenarios because checking and response implementation are directly linked such that checking always occurs before response implementation in real NPP operations. This link between checking and response implementation was maintained for scenario external validity. The three possible task sequence orders were:

- checking, response implementation, and detection

[4] Each student RO1 was partnered with a participant that had extensive experience with the scenarios and general study procedures. This experienced student served as the RO2 for the student group. The RO2 data were not analyzed.

- checking, detection, and response implementation
- detection, checking, and response implementation

Simulator Configurations

RIL 2020-11 (Volumes 1 and 2) contain 3 types of simulators. These are: a desktop, touchscreen, and hard panel simulator. All three simulator represented a Westinghouse pressurized water reactor. The desktop and touchscreen simulators used digitized analog controls, the hard panel used fully analog controls.

Desktop simulator

The full scope desktop simulator was comprised of a Dell 6.4GT/s, Intel XeonTM 5600 series processor, two 24" (16:10 aspect ratio) UXGA monitors with a total resolution of 3600 by 1200px, and a USB 3-button laser mouse with a scroll-wheel (Figure 3). The interaction design for the desktop interface required participants to use the mouse and scroll-wheel to view all the controls as not all the controls could fit in the display area of the desktop monitors. Participants had to use the mouse to activate the zoom feature (i.e., click on the "+" to zoom in and "-" to zoom out).

Figure 3. Desktop simulator

Touchscreen simulator

The full scope touchscreen interface is depicted in Figure 4. It consisted of eight 27-inch touchscreen Wide Quad High Definition (WQHD) monitor grids (two high by four wide) with a total resolution of 10240 by 2880px and had a touch-based interaction design. In RIL 2020-11, Volume 1, the touchscreen configuration was used by students. In RIL 2020-11 Volume 2, the touchscreen configuration was used by operators. The interface displayed the instrumentation and control panel in its entirety (i.e., removing the need for scrolling and zooming), but the large interface required participants to stand and move laterally to visually scan and interact with the interface (Figure 4). The only difference between the student and operator hardware set up was the screens were positioned on a table for the students, but they were mounted on a wall for operators.

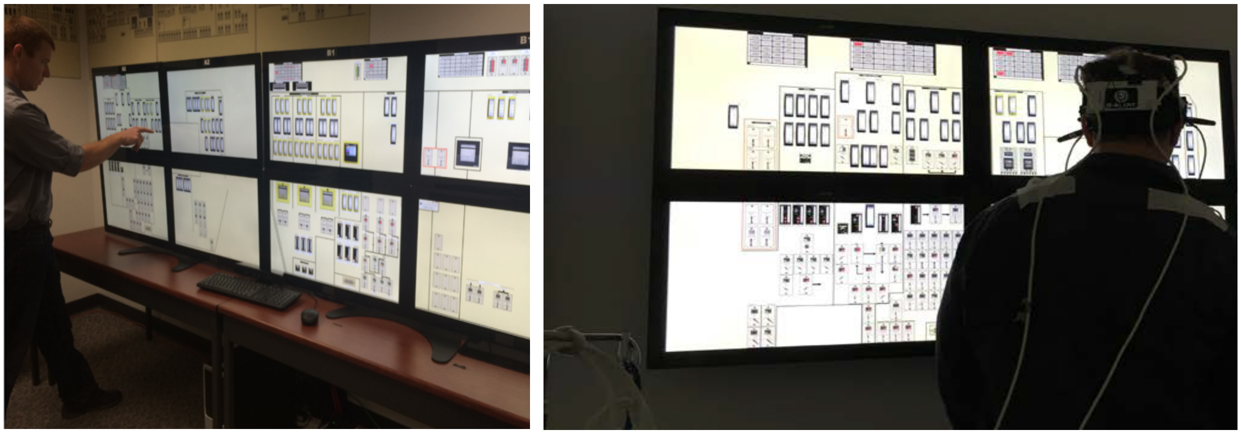


Figure 4. Touchscreen simulator. Left, student configuration, right operator configuration

Hard panel simulator

The hard panel simulator, like the two types of digital simulator was full scope. The hard panel simulator differed in that it was equipped with hard wired analog instrumentation and controls (I&C) which were fixed on conventional hard panels or bench boards. Operators were required to stand and move from panel to panel to manually manipulate the physical analog controls.

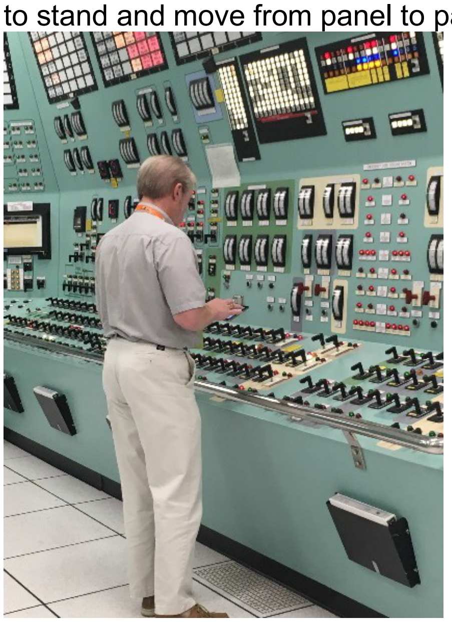


Figure 5. Former Operator Participants Interacting with the hard panel simulator.

Summary of RIL 2022-11, Study 1 Volume 1 Touchscreen Research with Students

This study was conducted to compare a desktop and touchscreen workstations in a student population. The aim was to understand performance and workload differences between the touchscreen and desktop HSIs and determine if student performance was comparable to operator performance. The desktop interface was smaller and could not accommodate a full view of the panels and therefore required the use of the mouse and keyboard to pan and zoom to access the controls required to execute the procedure.1 One-hundred fifty-two participants, ranging in age between 18 to 40 (85 men, 67 women), participated in this study.

The students used modified versions of the simulator panels to scale the cognitive difficulty of the checking, detection, and response implementation tasks to their level of expertise. The aim was to elicit performance from students (novices) that was in line with expert performance. Section 2.2.3 of the present report describes the differences between the panels used by students and those used by the operators.

The study confirmed that both task types (e.g., checking, detection, response implementation) and interaction modality (touchscreen, desktop) influenced multiple workload metrics and performance when using a touchscreen interface. The detection task, which required sustained monitoring of a small area of one of the panels, was generally the most difficult, relative to checking and response implementation. Additionally, student participants generally rated workload on the subjective measure lower than the objective measures suggesting differences between perceived and actual workload. These findings are discussed in detail in RIL 2022-11, Volume 1. The re-analysis will focus on a subset of the original measures and compare desktop to touchscreen performance, with an emphasis on the spatial position of the controls and potential ergonomic influences on performance and workload.

Summary of Study RIL 2022-11 Study 1, Volume 2: Touchscreen and Hard Panel Research with Former Operators

This study used a similar design to other HPTF research (see Appendix A), but rather than students, it included a sample of eighteen formerly licensed operators between the ages of 18 to 50 (14 men, 4 women). Participants had a range of operational experiences, including both Pressurized Water Reactors (PWRs) and Boiling Water Reactors (BWRs) in either commercial or naval nuclear power generation domains. Operators are experts and therefore did not require the reduction in the number of controls on the panel or a simplification in the control naming convention. The panels and control labels were reverted to their full complexity to scale task difficulty to the operators level of experience.

The study aimed to determine whether comparable effects observed in the student population generalized to an expert sample. Establishing equivalence between operators and non-operators for basic cognitive tasks associated with control room operations has been a core area for the HPTF. By establishing equivalence, it enables the use of non-expert populations as a stand-in for operators for skill- and rule-based tasks, enabling access to a greater number of potential participants.

The operators produced results in terms of workload and performance that were similar to students for the skill- and rule-based tasks that they were asked to perform (e.g., checking, detection, and response implementation, see section 2.2 for details). The operators reported significantly lower subjective workload ratings for the hard panel simulator compared to the touchscreen. However, this was reversed for the objective measures (performance, physiology) and these measures showed higher workload for the hard panel simulator and lower for the touchscreen. These results revealed broadly that workload was higher for the touchscreen than the hard panel simulator and that the perceived workload was not related to more objective measures (RIL 2020-11, Volume 2). These results also provided an initial validation that for skill- and rule-based tasks, students could be used as stand-ins for operators. This study also found differences between the RO1 and RO2 roles, which could be driven by differences in the spatial position of controls. This was one of the drivers for this ergonomics-focused re-analysis.

Interface Modifications

The experimental scenario (EOP-EPP-001) required each RO to interact with two control panels (A2 and C1 for RO1; A2 and B1 for RO2)[5]. To ensure comparable cognitive demands between the student and operator samples, the number of instrumentation and controls (I&Cs) in each panel used by the students was reduced systematically by 43%[6]. Additionally, I&C names in real control rooms often contain acronyms, which require knowledge of the system to understand. To focus on the skill-based tasks, the naming convention of specific I&Cs were also modified to eliminate the need for system knowledge for control identification and localization (for more information on the validation of the simplification approach see RIL 2022-11, Vol 1, Experiment 1). In the study conducted with former operators (RIL 2022-11, Vol 1, Experiment 2), the simplification modifications were reversed such that all simulator controls were presented. Figures 6, 7, and 8 show the panels used in the two studies referenced in this section.

A. Panel A2 Study 2 (simplified) **B. Panel A2 Study 3 (full complexity)**

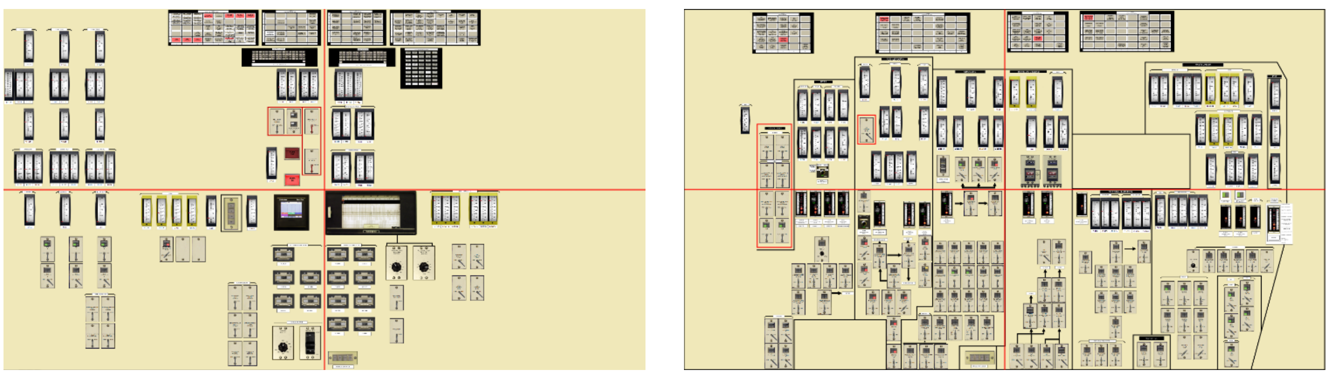


Figure 6. RO1 A2 panel.

[5] Only RO1 data is analyzed for the student sample. RO2 in the student sample was a highly trained student. Their presence was to enhance realism.

[6] Panel C1 is the simplest panel in the GPWR control room. This panel was used as the reference point to systematically simplify the environment for student participants while maintain realism and equivalent cognitive complexity for the experimental tasks. Further details on the scaling methodology can be found in RIL 2022-11, Volume 1.

A. Panel C1 Study 2 (simplified) **B. Panel C1 Study 2 (full complexity)**

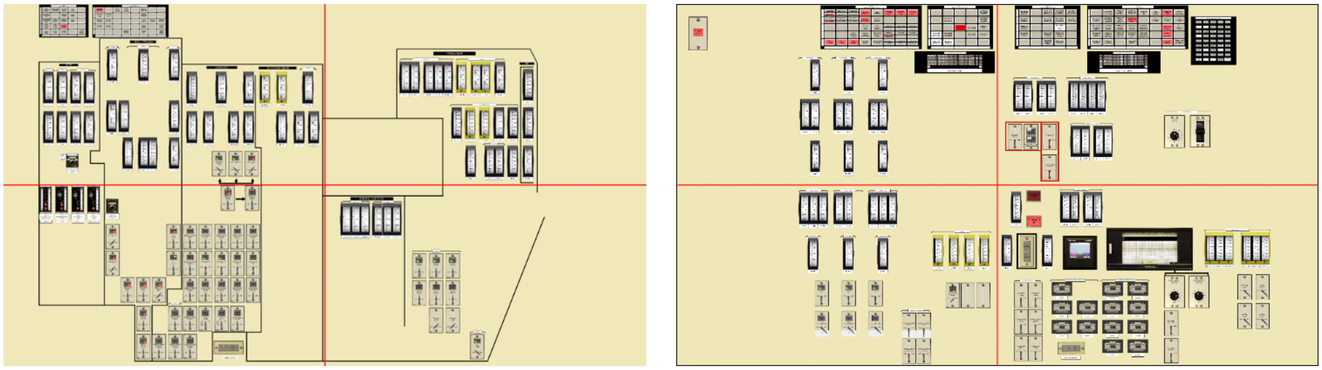


Figure 7. RO1 C1 panel

A. Panel B1 Study 2 (simplified) **B. Panel B1 Study 3 (full complexity)**

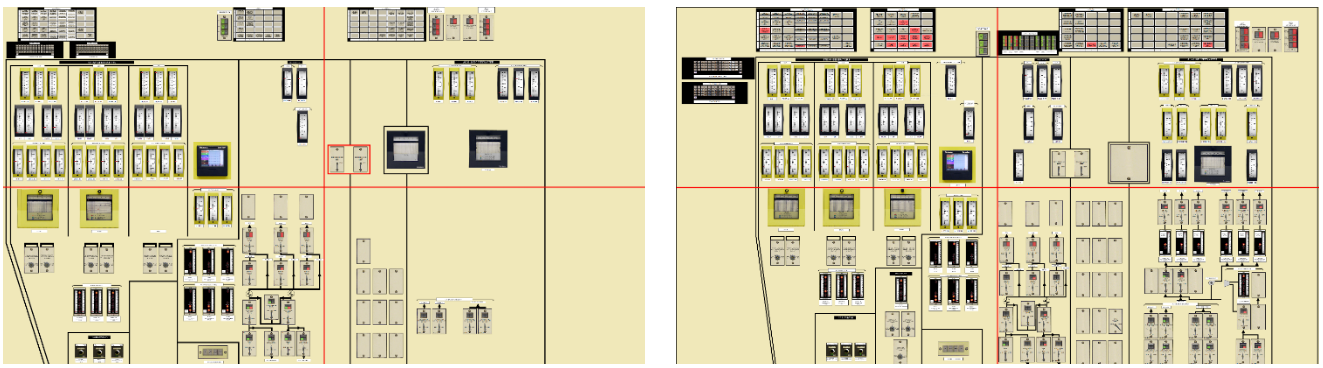


Figure 8. RO2 B1 panel

Task Type Re-grouping for Preliminary Ergonomic Analysis

To investigate operator performance as a function of the spatial location of the I&Cs, the I&Cs required for each of the three tasks were regrouped based on their vertical location on the panel. Since the original studies were not designed to examine the impact of spatial position on performance and workload, there is not a symmetrical distribution of controls across positions or task types. For example, for detection there is a split between the upper and lower panels for RO1, but for RO2 the controls are split between higher on the upper screen and lower on the upper screen. Tasks 1 and 3 and 2 and 4 were grouped together for each of the tasks. For example, two detection tasks (Step 2 and Step 4) for RO1 happened to be on one of the upper screens on A2 Panel and the other two detection tasks (Step 1 and Step 3) for RO1 appeared on one of the lower screens on C1 Panel. In this case, the location comparison was between upper and lower screens. In cases where all four steps of a task type happened to appear on only upper or lower screens the space was further divided into higher (50% from the top) and lower (50% from the bottom) sections. For example, for RO2 all four steps for the detection task were on the top screen in B1 Panel. Step 1 and Step 3 were grouped for analysis as they were on the higher section, and Steps 2 and 4 were grouped as they were on the lower section. The checking and response implementation tasks were grouped in a similar way for analysis.

Table 5. Task Location Summary

Task x Step	RO1 Screen Position	RO2 Screen Position
Detection	**RO1**	**RO2**
Step 1	Lower Screen	Higher on Upper Screen
Step 2	Upper Screen	Lower on Upper Screen
Step 3	Lower Screen	Higher on Upper Screen
Step 4	Upper Screen	Lower on Upper Screen
Checking	**RO1**	**RO2**
Step 1	Upper Screen	Higher on Lower Screen
Step 2	Lower Screen	Lower on Lower Screen
Step 3	Lower Screen	Lower on Lower Screen
Step 4	Lower Screen	Lower on Lower Screen
Response Implementation[7]	**RO1**	**RO2**
Step 1	Lower on Lower Screen (Clustered)	Lower on Lower Screen (Clustered)
Step 2	Lower on Lower Screen (Clustered)	Lower on Lower Screen (Clustered)
Step 3	Lower on Lower Screen (Clustered)	Higher on Lower Screen (Clustered)
Step 4	Higher on Lower Screen (Isolated)	Higher on Lower Screen (Clustered)

7 Due to a computer error in the original data collection period (2015), there is no performance-based data for response implementation available for re-analysis.

Dependent Measures

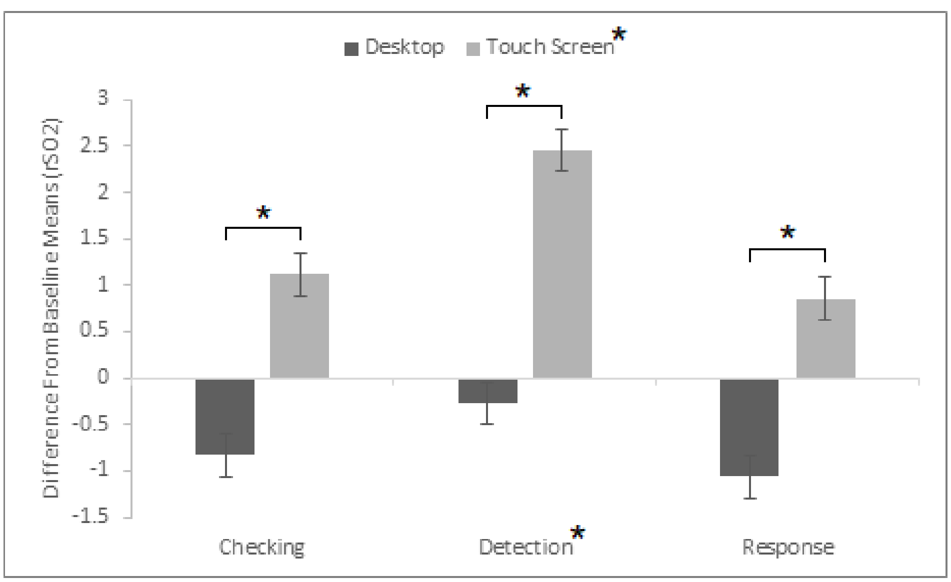
This re-analysis focuses on a subset of the dependent measures available from previous studies. Based on a review of the previous work and the ergonomics literature, the study team elected to focus on two performance measures, specifically the number of actions and response accuracy, the NASA-TLX (Hart & Staveland, 1988; Hart, 2006) survey for subjective workload, and heart rate variability (HRV)[8] and functional near-infrared spectroscopy (fNIRS) as physiological indices of workload. HRV measures the time differences between heart beats (inter-beat interval). Lower HRV is a general indication of higher workload. fNIRS is used to measure hemodynamic changes in oxygenated and deoxygenated hemoglobin in the prefrontal cortex. When the amount of oxygenated blood is higher, it is generally taken as an indication that workload is higher – greater oxygen resources are required when workload is higher. This multivariate assessment strategy is a core element of the HPTF approach and aids in strengthening confidence and interpretability in study results.

CHAPTER 3 RESULTS

Student Workload and Performance Using Touchscreens

Before presenting the re-analysis for the purpose of understanding potential touchscreen ergonomic issues, it is important to first understand if there is a difference between a desktop (i.e. keyboard and mouse) and touchscreen workstations.

In terms of physiology, there was higher rSO2, or regional cerebral oxygen saturation, when participants used the touchscreen compared to the desktop, suggesting that the touchscreen condition required a higher level of oxygen delivery to the brain (Figure 9). This can be associated with increased cognitive workload, as the brain needs more oxygen to perform complex tasks. The differences were particularly notable for the detection task, but there were also differences for checking and response implementation.



[8] Only available for students.

Figure 9. fNIRS Data for Students, Collapsed Across Hemisphere.

The HRV data did not align with the fNIRS data, based on the HRV data, students experienced greater workload while using desktop workstation compared to touchscreen.

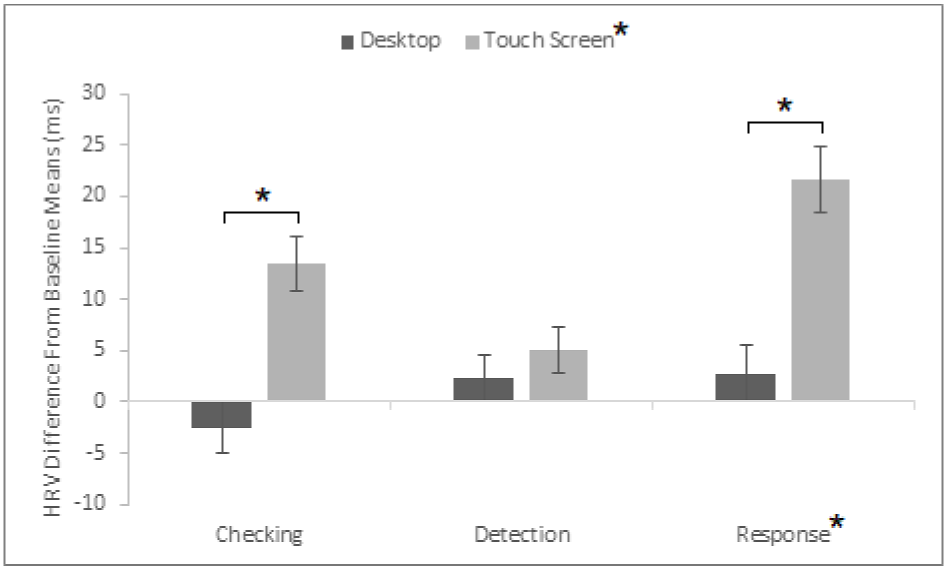


Figure 10. Heart Rate Variability Results for Students.

The NASA-TLX findings were somewhat contradictory to the fNIRS findings but consistent with the HRV data, students reported greater performance demand and effort for the desktop compared to the touchscreen. Higher perceived workload may have occurred on these particular subscales because of the amount of zooming and panning required to effectively interact with the panels on the desktop workstation.

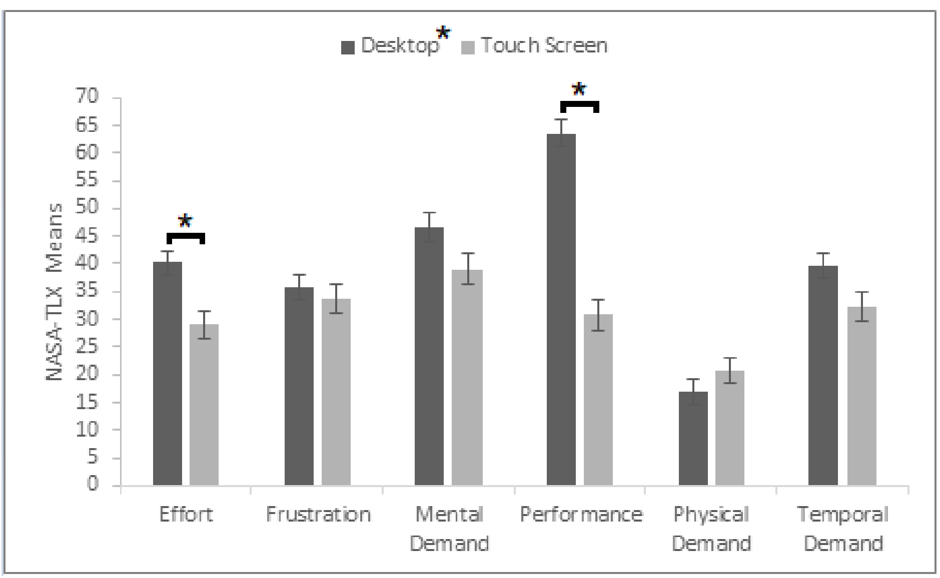


Figure 11. NASA-TLX Scores for Students Following Desktop and Touchscreen Use.

In terms of performance, students identified the correct controls significantly more often for the detection and response implementation tasks on the desktop workstation compared to the touchscreens. The differences likely only occurred in detection and response implementation because these two tasks required an active response to the information, whereas the response to checking was more passive (respond to a change and take action, versus log a response that information in the display is in the correct state).

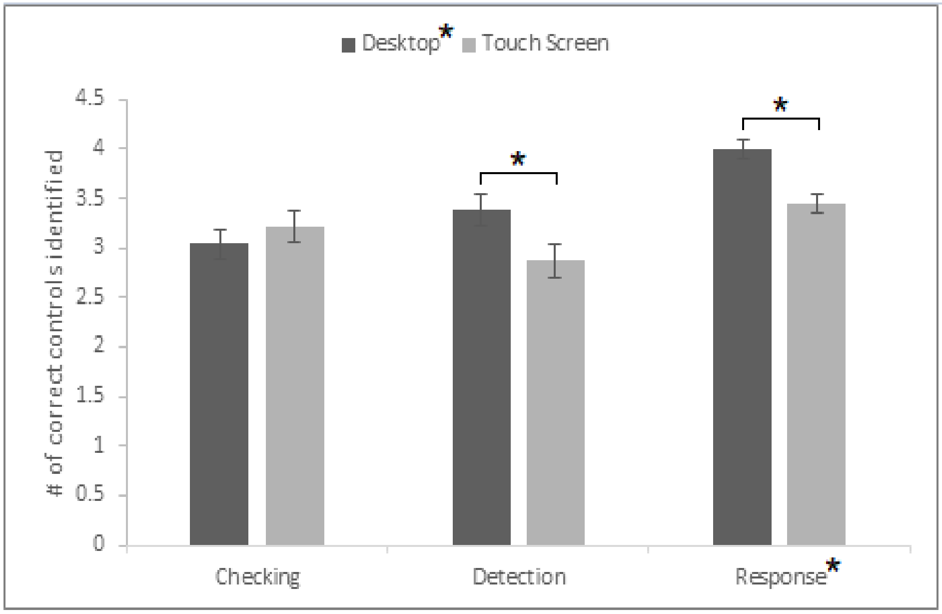


Figure 12. Performance Comparison between Keyboard and Mouse and Touchscreen.

A close look at the average number of actions for detection in the touchscreen group revealed that the position of the controls seemed to impact performance. There is an overall effect of procedure step driven by a statistically significant comparison between step one (lower) and two (upper) and step 1 (lower) and step four (upper). Referring back to Figure 3, the upper screens contained controls that occurred primarily at or near shoulder level, compared to the lower screen which required the participant to view the control from the top-down, potentially introducing a parallax error (see Table 3) or creating a condition where poorer hand and/or arm control occurred because of the requirement that participants bend to view and engage the controls on the lower screens.

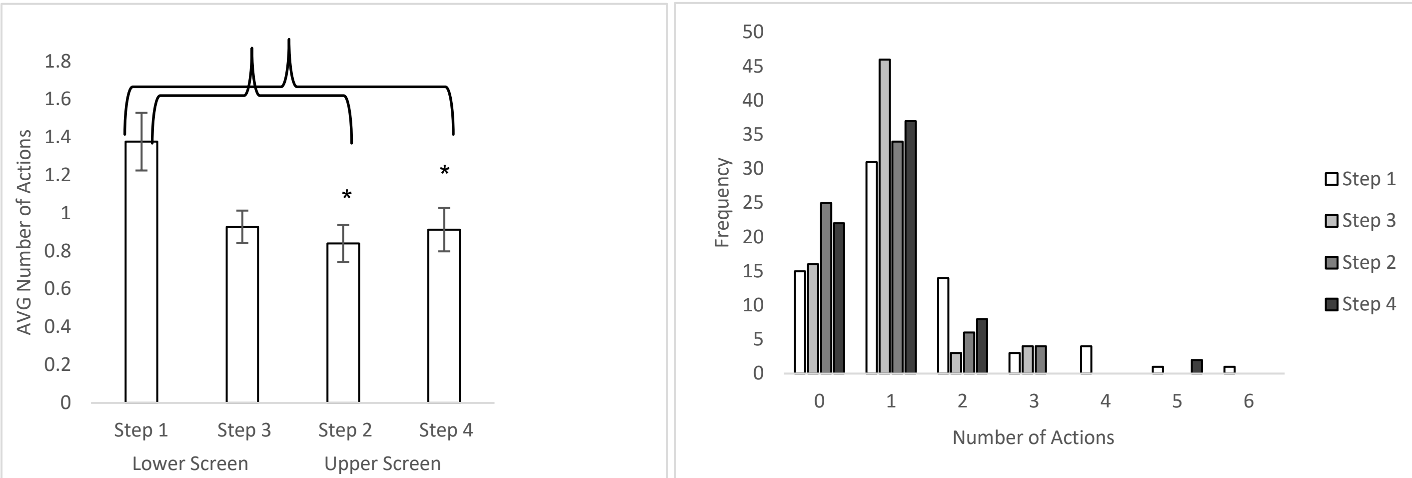


Figure 13. Average Number of Actions as a Function of Control Position.

<u>Operator Data: Workload and Performance While Using Touchscreen</u>

Physiological Workload

fNIRS

The fNIRS sensor was only available for data collection on RO1. As with the students, there was significantly higher rSO2, or regional cerebral oxygen saturation, when participants used the touchscreen compared to the desktop, suggesting that the touchscreen condition required more cognitive effort than the hard panel (analog) simulator. This difference between simulators was largest and statistically significant for the detection task. It is likely that the other tasks did not reach significance because of the small sample size, high variability, and limited number of trials.

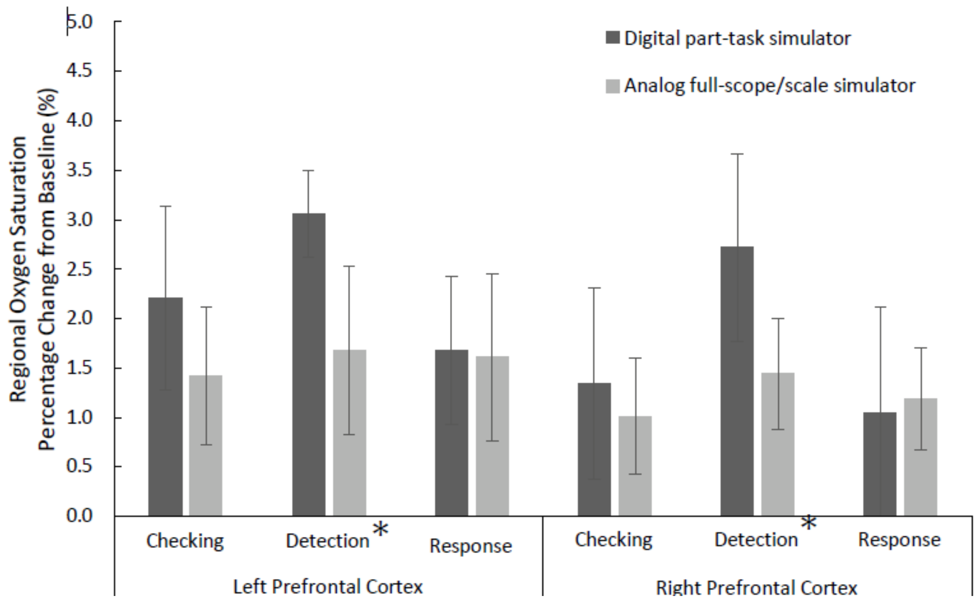


Figure 14. Cognitive Workload Demonstrated by rSO2.

Subjective Workload

The NASA-TLX results for operators were mixed between the balance of plant operator (BOP) and RO. The RO tended to rate the hard panel (analog) simulator as inducing more workload than the BOP, who tended to rate the digital (touchscreen) simulator as higher workload. The RO experienced greater workload for the mental, temporal, and performance subscales. The BOP experienced greater workload for frustration and effort subscales. These differences make sense in the context of the two simulator types. The hard panel simulator required participants to use a greater variety of manual gestures and navigate a larger physical space. These differences could induce additional mental, temporal, and performance demands. The BOP would have experienced the same challenges on the hard panel simulator, however, they generally rated workload as equivalent on those subscales and instead reported higher frustration and effort. It is possible that these workload metrics point to the touchscreen being difficult to use for the BOP participants.

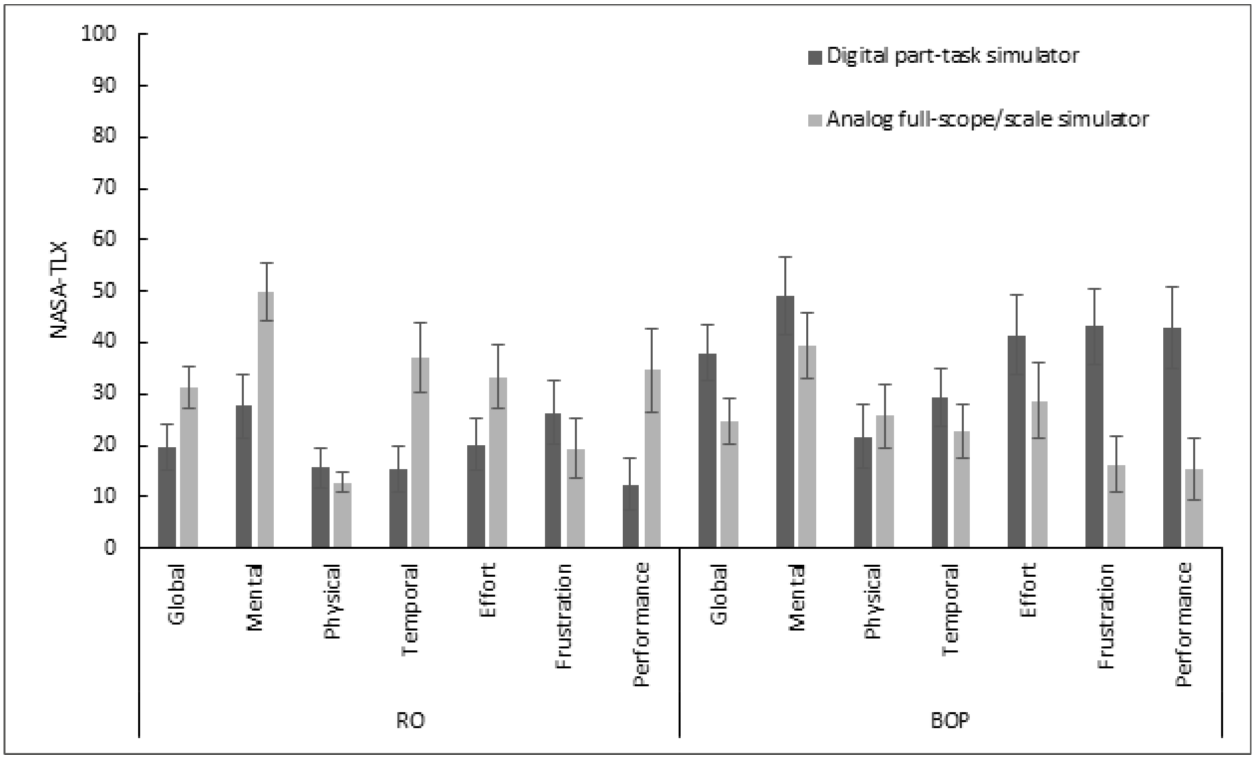


Figure 15. Comparison of NASA-TLX Scores for Touchscreen and Hard Panel Simulators.

Performance

Task performance measures were only available for operators for the digital (touchscreen) simulator group. As with the students, the focus of the re-analysis is on detection, specifically the number of actions. The performance analysis for operators focuses only on the spatial position effects. Table 3 outlines all the possible spatial configurations for the controls used by the operators. There is only one noteworthy significant finding in the reanalysis for spatial positional effects. The overall effect of the number of actions at each step level trended towards significance. There were a greater number of actions attempted on the upper high screens, which would have been above the operators' shoulder level than the upper lower screen, which would have been at or slightly below the operators' shoulder level.

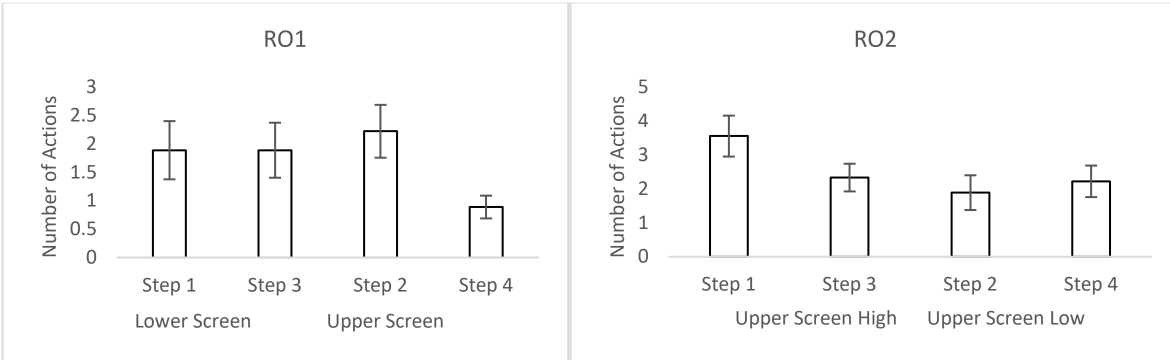


Figure 16. Average Number of Actions Attempted for RO1 and RO2 at Different Touchscreen Spatial Positions.

CHAPTER 4 DISCUSSION AND CONCLUSIONS

The summarized results from previous HPTF simulator studies suggest that there may be human factors/ergonomics challenges associated with touchscreen use. These results were revealed in greater detail when performance was analyzed in terms of the spatial position of controls. For example, the participants' performance was worse for some areas (e.g., I&Cs below the shoulder level) than others. There was also evidence for variation in perceived and actual workload across the different modalities (desktop workstation, glass panel, hard panel).

Is there evidence of differences in workload and performance between keyboard and mouse, hard panels, and touchscreen controls?

Students reported higher workload on the performance and effort subscales of the NASA-TLX for the desktop condition compared to the touchscreen condition, and these ratings were consistent with the HRV measure of workload. However, the fNIRS and performance data paint the opposite picture – workload is higher and response efficacy lower (e.g., more action attempts) when using the touchscreen compared to a traditional keyboard and mouse. The difference in workload between the two interaction modalities could be driven by the interaction design in the desktop condition. Navigating the digitalized panels on a desktop computer required significant panning and zooming, which may influence participants self-rating for workload and their HRV.

Operators reported mixed effects for workload comparisons between the touchscreen and the hard panel simulator. It is possible these effects varied because of the small operator sample and variation in spatial position of the controls, which varied between the BOP and RO1 and between the touchscreen and the hard panel simulator. The fNIRS and performance data for operators demonstrate the same pattern as the students – higher processing resource (rSO2) demands and poorer response execution in the touchscreen condition.

Taken together the results suggest that there may be higher workload associated with the use of touchscreens, however, more research is needed. Additionally, the variation in workload pattern depending on the measurement technique reinforces that a multivariate approach can reveal more nuanced trends in workload changes compared to single measure or subjective measure-based techniques.

Is there evidence of differences based on the spatial position of controls?

The student detection data was re-analyzed based on spatial position of the controls. There was a significant difference between the controls placed above and below the shoulder (upper vs. lower screens). There was also a trend for the students to produce a greater number of actions on the lower screens, suggesting usability challenges for those spatial positions. These differences could be driven by a parallax error (Orphanides and Nam, 2017) or the participant's body size. The performance differences between spatial position of controls for operators were non-significant, likely due to a small sample size in the operator group.

The operator data did trend towards actions requiring fewer attempts on the lower portion of the upper screen (shoulder height) compared to the higher portion of the upper screen (above shoulder height). These upper screen errors could also be caused by parallax - a special subtype of visual targeting error where, because of the perceived angle of approach, the user misses the target, usually touching too low (see Figure 2, Table 3, and Orphanides & Nam, 2017). Occlusion could also contribute to errors when interacting with controls above shoulder height, as the participants hand may block the control when approaching it from below the control position.

There is substantial research available discussing touchscreen interaction design broadly. However, there is very limited work focused on spatial positional effects and even less discussing touchscreens in a nuclear context. Kang and Shin (2017) found that participants responded faster to information in lower portions of a touchscreen when that device was positioned as a desktop workstation. Similarly, an ergonomics study in the aviation domain reported that pilots' physical effort varies depending on the location of touchscreen and task duration (Barbé et al., 2012). Chourasia et al., (2013) found that body position had a significant impact on performance when touchscreen buttons were smaller in size. Specifically, when participants had to respond from a standing instead of a seated position, they made more errors (misses) and produced slower reaction times. These previous studies are consistent with the overall pattern of results from the re-analysis – spatial position matters and can influence both workload and performance.

Height as an influencing factor to performance

The effect of I&C location on performance in the detection task confirmed that I&C locations, more specifically, the vertical height, is a potential influencing factor to the task performance in NPP operations. Comparing performance in the tasks associated with I&Cs on the upper screens versus I&Cs on the lower screen, participants from both student novice and former operator samples had to make fewer attempts to complete an action on the screen or portion of the screen that was closest to shoulder height. While reaching from the shoulder is an unsupported mid-air gesture and less control and more fatigue would be expected, the specific touchscreen set ups used in the studies analyzed here would have had information positioned either so low that it required bending or so high that it required an above the shoulder reach (also an unsupported mid-air gesture). The shoulder height reach would have been the easiest to control from an ergonomic perspective. Future research with more observations at different spatial positions is needed to better understand the positional effects. Additionally, future research should capture participant height as a metric to better characterize the severity of the potential fatigue and loss of gestural control at extreme upper and lower positions.

Implications and Limitations

Connection to Guidance

The NRC *Human-System Interface Design Review Guidelines* (NUREG-0700, Revision 3) listed several advantages of using touchscreen devices, such as no separate input device needed, programmable interface, fast access, direct manipulation of targets, input/output in same location, intuitive, natural pointing action, and less training required. However, results from the re-analysis of data from previous studies suggest that I&C locations can be an influencing factor to the performance associated with the I&C. ISO 9241-303 indicates that for a typical working environment with an approximately vertical position of the upper body, the work place and the visual display should permit the user to view the screen with a gaze angle from 0° to 40° and a head-tilt angle of from 0° to 25°. In our experimental setting, the lower section of the upper screens was approximately shoulder level for most participants. This positioning led to some controls being easy to interact with and some requiring more effort and being more error prone, suggesting that the height of the screen may need to be adjustable for individual operator's needs.

Currently, NUREG-0700 treats guidance related to touchscreens like other digital input devices, however, a review of the guidance in NUREG-0700 Section 11.2, "Workstations Containing Primarily Analog HSIs" reveals design guidance that may generalize to touchscreens. For example, Section 11.2.1.1-1, Table 11.4 provides information related to console height for visibility. Section 11.2.1.1-4 describes the minimum distance a control should be from the front edge of the console to prevent accidental activation, a known safety issue across human factors domains. Section 11.2.1.2 provides guidelines for standup workstations that may be applicable for touchscreen workstations if a control room with touchscreens elects to maintain the "bay" style configuration. Finally, Section 11.2.2.3.1 details the amount of separation needed between controls to prevent one control from interfering with another. While this may be more of an issue with switches bumping into one another in an analog control system; in a touchscreen system, this could be an issue of an operator's hand accidentally opening a valve while they rest their hand on the display while opening an adjacent valve.

Limitations

While the findings of this analysis are useful and consistent with the existing literature, the objective of the original studies was not to focus on ergonomics issues and, therefore, there are some limits in the interpretation of these results. First, the location of I&Cs is not a manipulated factor in the original experimental design. Although the I&Cs are grouped for analysis using the positions shown on the screen, the vertical height from floor is not measured. In addition, the lack of participants' height information makes it impossible to take the individual differences in height into consideration. Second, some of the important performance metrics, such as reaction time and number of errors made were not available for the checking and response implementation tasks, limiting the scope of the re-analysis of performance to only detection. The scope of the workload analysis was broader but still limited by data availability. It is possible that other tasks could produce different patterns of results. For example, response implementation requires a wider range of gesture types than detection. It is possible the differences between the touchscreen and other interaction modalities would be more pronounced in response implementation than detection. Third, due to the availability of former operators for participating in the studies, the operator samples were much smaller than the students. The small sample size limited the statistical power of the analyses of operator data.

Future Studies

This re-analysis presented preliminary indications that there may be limitations inherent to the touchscreen modality that limit their usefulness in a training, licensing exam, or validation and verification context. When a plant is making choices about their simulator technology and they opt for a non-plant referenced simulator, they will likely choose to match the plant as closely as possible. While this seems like the right course of action on the surface, applying this “maximize fidelity” approach could create a glass top simulator with significant limitations. A restarting plant may have a substantial number of analog controls that require complex manual gestures to operate effectively. If the plant tries to match the physical fidelity as closely as maximally possible, they may be creating a situation where operators make errors because the same human factors limitation of the plant are replicated in the glass top and those limitations are exacerbated by the conversion from 3D to 2D controls. This may be particularly apparent when controls are too close together, or the gesture required to engage the control (turn a handle or open a valve) is difficult to execute.

The results of this report informed the development of a follow-up study making a direct comparison between a hard panel and glass top simulator using operators as participants performing complex scenarios and using simulator-based job performance measures. The scenarios will include a broader range of tasks and gesture types, including known complex gestures, such as turning handles, opening valves with long stroke times, manual rod control, and manual turbine and steam generator control. This will provide more opportunities to observe differences between the glass top and hard panel simulator. Additionally, the study will include a formal ergonomic evaluation to separate out the impact of physical and task factors on operator performance.

Conclusions

The modernization of conventional large light water reactors, as well as newly built large light water, small modular, and other advanced reactor concepts will be an opportunity to introduce new digital HSIs into the control room and the training simulator. Touchscreens could very well be part of this new digital landscape. As such, it is important to understand the potential for performance and ergonomics issues associated with touchscreen interfaces.

Understanding how novel HSIs might change operator performance on common control room tasks and documenting the impact those HSIs have on workload is an important first step in developing guidance for those HSIs. NUREG-0700, Revision 3 already has guidance on touchscreens, however, it may be that future guidance updates may want to borrow from analog control ergonomics sections of NUREG-0700 or include new guidance entirely, particularly if designers elect to maintain the same control room form factor and design touchscreen “bays”. Basic research like this study provides an early indication of where guidance updates may be needed.

The HPTF methodology of systematically evaluating new technologies and their impact on workload can also aid in the collection of new HRA data. In the Integrated Human Event Analysis System for Event and Condition Assessment (IDHEAS-ECA; Xing, Chang, & DeJesus, 2020), cognitive failure modes (CFMs) are used to model the failure of a critical task. The human error probability estimate of a CFM is based on the assessment of one or more performance influencing factors (PIFs). These PIFs include things such as the HSI attributes and task factors that contribute to workload (multitasking, mental fatigue, time pressure, task familiarity, physical demands). A systematic and multidimensional assessment of performance and workload during a scenario enables data to be collected under conditions with a greater degree of both confidence and control.

CHAPTER 5 REFERENCES

Barbé, J., Chatrenet, N., Mollard, R., Wolff, M., & Bérard, P. (2012). Physical ergonomics approach for touch screen interaction in an aircraft cockpit. *Proceedings of the 2012 Conference on Ergonomie et Interaction Homme-Machine*, 9–16.

Boring, R. L. (2012). *Fifty years of THERP and human reliability analysis* (No. INL/CON-12-25623). Idaho Falls, ID: Idaho National Laboratory (INL).

Breuninger, J. (2020). *Suitability of touch gestures and virtual physics in touch screen user interfaces for critical tasks* (Doctoral dissertation, Technische Universität München).

Chourasia, A. O., Wiegmann, D. A., Chen, K. B., Irwin, C. B., & Sesto, M. E. (2013). Effect of sitting or standing on touch screen performance and touch characteristics. *Human Factors*, *55*(4), 789–802.

Davis, K. G., Hammer, M. J., Kotowski, S. E., & Bhattacharya, A. (2014). An ergonomic comparison of data entry work using a keyboard vs. touch screen input device while standing and sitting. *J Ergonomics S*, *4*(2).

Hansberger, J. T., Peng, C., Mathis, S. L., Areyur Shanthakumar, V., Meacham, S. C., Cao, L., & Blakely, V. R. (2017, May). Dispelling the gorilla arm syndrome: the viability of prolonged gesture interactions. In *International conference on virtual, augmented and mixed reality* (pp. 505-520). Cham: Springer International Publishing.

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

Hart, S. G. (2006, October). NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 50, No. 9, pp. 904-908). Sage CA: Los Angeles, CA: Sage publications.

Hughes, N., D'Agostino, A., & Reinerman-Jones, L. (2017). The NRC's Human Performance Test Facility: Methodological considerations for developing a research program for systematic data collection using an NPP simulator. *Proceedings of the Enlarged Halden Programme Group (EHPG) meeting*, September 24-18, 2017, Lillehammer, Norway.

Hughes, N., D'Agostino, Dickerson, K., Matthews, G., Reinerman-Jones, L, Barber, D., Mercado, J., Harris, J., & Lin, J. (2022). *Human Performance Test Facility (HPTF): Systematic Human Performance Data Collection Using Nuclear Power Plant Simulator: A Methodology - Volume 1.* (RIL-2022-11). Washington, D.C.: U.S. Nuclear Regulatory Commission.

Hughes, N**.**, D'Agostino, Dickerson, K., Matthews, G., Reinerman-Jones, L, Barber, D., Mercado, J., Harris, J., & Lin, J. (2022). *Human Performance Test Facility (HPTF): Comparing Operator Workload and Performance Between Digitized and Analog Simulated Environments - Volume 2. (RIL-2022-11*). Washington, D.C.: U.S. Nuclear Regulatory Commission.

Hughes, N., Dickerson, K., Lin. J., Matthews, G. & Barber, D. (2022). *Human Performance Test Facility (HPTF) - Supplementary Exploratory Analyses of Sensitivity of Workload Measures – Volume 3 (RIL-2022-11).* Washington, D.C.: U.S. Nuclear Regulatory Commission.

International Organization for Standardization. (2011). Ergonomics of human-system interaction — Requirements for electronic visual displays (ISO Standard No. 9241:303).

Kang, H., & Shin, G. (2017). Effects of touch target location on performance and physical demands of computer touchscreen use. *Applied Ergonomics*, *61*, 159–167.

Li, W. C., Liang, Y. H., Korek, W. T., & Lin, J. J. (2022, June). Assessments on human-computer interaction using touchscreen as control inputs in flight operations. In *International Conference on Human-Computer Interaction* (pp. 326-338). Cham: Springer International Publishing.

Mosleh, A., & Chang, Y. H. (2004). Model-based human reliability analysis: Prospects and requirements. *Reliability Engineering & System Safety*, *83*, 241-253.

Nagarajan, K., Silva, A., & Lee, H. (2020). Insights for touchscreen interface guidelines for medical devices: A case study of a usability test.

NRC (2016). *Standard Review Plan* (NUREG-0800); Chapter 18, Revision 3, Human Factors Engineering. Washington, D.C.: U.S. Nuclear Regulatory Commission. Retrieved from: https://www.nrc.gov/ reading-rm/ doc-collections/nuregs/staff /sr0800/ch18/

NRC (2019). *Human-System Interface Design Review Guidelines* (NUREG-0700, Revision 3). Washington, D.C.: U.S. Nuclear Regulatory Commission. Retrieved from: https://www.nrc.gov/reading-rm/doc-collections/ nuregs/ staff/sr0700/

NRC (2012). *Human Factors Engineering Program Review Model* (NUREG-0711, Revision 3). Washington, D.C.: U.S. Nuclear Regulatory Commission. Retrieved from: https://www.nrc.gov/reading-rm/doc- collections/nuregs/staff/sr0711/

O'Hara, J. M., & Higgins, J. C. (2010). *Human-system interfaces to automatic systems: Review guidance and technical bases* (BNL-91017-2010). *Human factors of advanced reactors (NRC JCN Y-6529).* Washington, DC: United States Nuclear Regulatory Commission.

O'Hara, J. M., Higgins, J. C., Brown, W. S., Fink, R., Persensky, J., Lewis, P., & Boggi, M. (2008). *Human factors considerations with respect to emerging technology in nuclear power plants* (NUREG/CR-6947). Washington, DC: United States Nuclear Regulatory Commission.

Orphanides, A. K., & Nam, C. S. (2017). Touchscreen interfaces in context: A systematic review of research into touchscreens across settings, populations, and implementations. *Applied ergonomics*, *61*, 116-143.

Reinerman-Jones, L. E., Guznov, S., Mercado, J., & D'Agostino, A. (2013). Developing methodology for experimentation using a nuclear power plant simulator. In D. D. Schmorrow, & C. M. Fidopiastis (Eds.), *Foundations of augmented cognition* (pp. 181-188). Heidelberg, Germany: Springer.

Reinerman-Jones, L. E., Lin, J., Matthews, G., Barber, D., & Hughes, N. (2019). *Human performance test facility experiment 4: Former operator workload and performance comparison between two simulated environments.* Rockville, MD: United States Nuclear Regulatory Commission.

Reinerman-Jones, L. E., Matthews, G., Harris, J., Barber, D., Hughes, N., & D'Agostino, A. (2018). *Human performance test facility experiment 3: Former operator workload and performance on three tasks in a simulated environment.* Rockville, MD: United States Nuclear Regulatory Commission.

Reinerman-Jones, L. E., & Mercado, J. (2014). *Human performance test facility task order 1 technical report* (JCN # V621). Rockville, MD: United States Nuclear Regulatory Commission.

Reinerman-Jones, L. E., Teo, G., & Harris, J. (2016). *Human performance test facility task order 1 technical report* (JCN # V621). Rockville, MD: United States Nuclear Regulatory Commission.

Swain, A. D., & Guttmann, H. E. (1983). *Handbook of human-reliability analysis with emphasis on nuclear power plant applications. Final report* (No. NUREG/CR--1278). Albuquerque, NM: Sandia National Labs.

Xing, J., Chang, Y. J., & DeJesus, J. (2020). *Integrated human event analysis system for event and condition assessment* (IDHEAS-ECA).(NUREG-2256) US Nuclear Regulatory Commission, Washington, DC.

APPENDIX A

Statistics for Reported Results

RIL 2022-11, Volume 1 Students

fNIRS (Functional Near-Infrared Spectroscopy)

A 3 (checking, detection, and response implementation) × 2 (left and right hemisphere) × 2 (interface type: desktop and touchscreen) mixed ANOVA was run to determine if oxygenation was differed significantly across task types and between interface types. A significant main effect was found for task type, $F(1.76, 252.27) = 49.10$, $p< .001$, $\eta_p^2 = .256$, such that the detection task ($M = 1.09$) resulted in a greater increase in blood oxygenation from baseline compared to the checking ($M = 0.14$) and response implementation ($M = -0.10$) tasks.

A significant main effect was also found for interface type, $F(1, 143) = 57.72$, $p< .001$, $\eta_p^2 = .29$. The touchscreen interface group ($M = 1.477$) experienced a greater increase in blood oxygenation from baseline compared to the desktop interface group ($M = - 0.72$). There was a significant main effect for hemisphere, $F(1, 143) = 11.26$, $p = .001$, $\eta_p^2 = .07$, the rise in oxygenation from baseline was higher overall in the right hemisphere ($M = 0.57$) compared to the left hemisphere ($M = 0.19$), and this was true across the different interface and task types.

In addition, a significant interaction between task type and interface type was found, $F(1.76, 252.27) = 6.53$, $p = .003$, $\eta_p^2 = .044$. The touchscreen interface group experienced a significant increase in blood oxygenation from baseline, whereas the desktop interface group experienced a significant decrease in blood oxygenation from baseline. These differences were most pronounced in the detection task.

NASA-TLX

A 3 (task type: checking, detection, and response implementation) × 2 (interface type: desktop and touchscreen) mixed ANOVA was conducted for each of the NASA-TLX subscales. Task type was a repeated-measures factor, and the interface type was a between-subjects variable. The ANOVAs were used to determine if there was a significant workload difference between task types, interface types, and if there were overall differences in the ratings across the subscales. The analysis would also reveal if task type effects differed for the two types of interfaces, and if different combinations of task and interfaces elicited different patterns of workload response, as tapped by the NASA-TLX subscales.

A significant main effect was found for task type, $F(2, 296) = 9.66$, $p < .001$, $\eta_p^2 = .06$, such that, in general, participants experienced greater workload during the detection task type ($M = 38.76$) compared to the checking ($M = 34.30$) and response implementation ($M = 34.04$) task types.

In addition, a significant main effect was found for the sub-scales of the NASA-TLX, $F(3.07, 453.95) = 50.89$, $p < .001$, $\eta_p^2 = .256$, such that overall, participant reported higher ratings on the performance (M = 47.152) and mental demand (M = 42.836) subscales compared to the other subscales.

Examining the effects of task on the subscales, showed a significant interaction effect between the task types and sub-scales on the NASA-TLX, $F(6.71, 992.36) = 19.50$, $p < .001$, $\eta_p^2 = .12$. Not only did the detection task induce the highest amount of workload overall, but it appears that the increase was especially marked for frustration workload.

Furthermore, a significant main effect for interface type was found, $F(1, 148) = 15.56$, $p < .001$, $\eta_p^2 = .095$, such that workload ratings were generally higher for the desktop interface (M = 40.46) compared to touchscreen interface (M = 30.93) groups. This increase in workload in the desktop interface group was much greater for the performance and effort subscales, as reflected in the significant interaction effect between the sub-scales on the NASA-TLX and interface types, $F(3.07, 453.95) = 21.30$, $p < .001$, $\eta_p^2 = .126$.

Electrocardiogram (ECG) – Heart Rate Variability (HRV)

A significant main effect was found for task type, $F(1.79, 254.09) = 13.79$, $p < .001$, $\eta_p^2 = .09$. The response implementation task (M = 12.21) resulted in significantly greater increases from baseline compared to the checking (M = 5.46) and detection (M = 3.65) task types. A significant main effect was also found for interface type, $F(1,142) = 14.55$, $p < .001$, $\eta_p^2 = .09$. Participants that used the touchscreen interface (M = 13.37) experienced greater increases in HRV from baseline compared to participants that used the desktop interface (M = 0.84). There was a significant interaction effect between task and interface type, $F(1.79, 254.09) = 12.48$, $p < .001$, $\eta_p^2 = .08$, in which the largest differences between the interface groups were found for the checking and response implementation task.

Performance – Number of Correct Controls

Navigation variables included: locating a correct control, the number of additional attempts to locate a correct control and locating a correct control on the first attempt. Three 3 (task type: checking, detection, and response implementation) × 2 (interface type: desktop and touchscreen) mixed ANOVAs were conducted for each of the three measures to determine if there was a significant difference between task types and between interface types. The analyses also revealed if the two interface groups showed similar patterns of differences in performance across the tasks. Task type was a repeated-measures variable and interface type was a between-subjects variable.

For locating a correct control, there was a significant main effect for task type, $F(1.79, 264.91) = 12.80$, $p < .000$, $\eta_p^2 = .08$, such that participants were able to correctly locate more controls for the response implementation task (M = 3.72) compared to the checking (M = 3.13) and detection (M = 3.12) task types. A significant main effect was found for interface type, $F(1, 148) = 6.25$, $p = .014$, $\eta_p^2 = .04$. Participants were able to correctly locate more controls using the desktop interface (M = 3.47) compared to the touchscreen interface (M = 3.18). A significant interaction effect was found between task type and interface type, $F(1.79, 264.91) = 4.51$, $p = .012$, $\eta_p^2 = .03$. Participants using the desktop interface were able to correctly locate more controls during the detection and response implementation tasks, but not in the checking task.

For the additional attempts to locate a correct control, a significant main effect was found for task type, $F(1.01, 148.52) = 8.94$, $p < .001$, $\eta_p^2 = .06$. The detection task (M = 6.66) required significantly more attempts to locate a correct control compared to the checking (M = 0.23) and response implementation (M = 0.54) task types. A significant main effect was found for interface type, $F(1, 148) = 4.22$, $p = .042$, $\eta_p^2 = .03$. Participants using the touchscreen interface (M = 3.914) required significantly more attempts to locate a correct control compared to participants using the desktop interface (*M* = 1.04). A significant interaction effect was found between task type and interface type, $F(1.01, 148.52) = 4.14$, $p = .044$, $\eta_p^2 = .03$. The differences between the touchscreen and desktop groups were most prominent during the detection task.

Students – New Analysis – RIL 2022-11, Volume 5 (This Document)

Performance – Number of Correct Actions as a Function of Spatial Position

A close look at the average number of actions for detection in the touchscreen group revealed that the position of the controls seemed to impact performance. A repeated measures ANOVA for 4 levels of procedure step [lower 1 and 3; upper 2, 4] revealed an overall effect of procedure step $F(3, 204) = 5.76$, $p < .001$, $\eta_p^2 = .08$. This effect was driven by statistically significant comparisons between procedure steps one (lower) and two (upper), $F(1, 68) = 11.13$, $p = .001$, $\eta_p^2 = .14$) and steps 1 (lower) and four (upper), $F(1,68) = 6.59$, $p = .012$, $\eta_p^2 = .08$).

Operators RIL 2022-11, Volume 2

fNIRS (Functional Near-Infrared Spectroscopy)

Due to the restriction of the sensor connection, the fNIRS sensor was only available for participants in the RO role. A 3 (task type: checking, detection, and response implementation) × 2 (hemisphere: left and right) × 2 (simulator: digital, part-task and analog, full-scope/scale) mixed ANOVA was run to determine the effect of task types and simulator types on regional oxygen saturation (rSO2) for the left and right prefrontal cortex. There was a main effect for task type, $F(2,32) = 5.48$, $p < .01$, $\eta_p^2 = .26$. Compared to the response implementation task type (*M* = 1.38), the detection task type (*M* = 2.22) showed a greater increase from baseline.

NASA-TLX

A series of 2 (simulator: digital, part-task and analog, full-scope/scale) × 2 (operator Role: RO and BOP) ANOVAs were run for each of the subscales of the NASA-TLX. A significant main effect of simulator was found for frustration. Participants in the digital, part-task simulator condition (*M* = 34.72) reported higher frustration ratings than participants in the analog, full-scope/scale simulator condition (*M* = 17.83), $F(1,44) = .12$, $p < .05$, $\eta_p^2 = .14$. There was no significant main effect of operator role for any of the subscales in the NASA-TLX.

A similar pattern of significant interactions between simulator type and operator role were found for global workload, $F(1,44) = 6.80$, $p < .05$, $\eta_p^2 = .13$; mental demand, $F(1,44) = 5.65$, $p < .05$, $\eta_p^2 = .11$; temporal demand, $F(1,44) = 5.38$, $p < .05$, $\eta_p^2 = .11$; and performance, $F(1,44) = 10.75$, $p < .01$, $\eta_p^2 = .20$. In the digital, part-task simulator condition, the BOP perceived higher workload and reported better performance, whereas in the analog, full-scope/scale simulator condition, the RO reported higher workload and better performance.

Operators – New Analysis – RIL 2022-11, Volume 5 (This Document)

Performance – Number of Actions as a Function of Spatial Position

The overall effect of the number of actions at each step level trended towards significance. There were a greater number of actions attempted on the upper high screens, which would have been above the operators' shoulder level than the upper lower screen, which would have been at or slightly below the operators' shoulder level.

$F(3,48) = 2.345$, $\eta_p^2 = .13$ $p = .08$ and a significant effect for the comparison of scenario steps 1 and 2 for RO2 $F(1,16) = 4.235$, $\eta_p^2 = .21$ $p = .05$.