U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

RIL 2022-11

HUMAN PERFORMANCE TEST FACILITY (HPTF)

VOLUME 2 – COMPARING OPERATOR WORKLOAD AND PERFORMANCE BETWEEN DIGITIZED AND ANALOG SIMULATED ENVIRONMENTS

Date Published: January, 2023

Prepared by:
N. Hughes[1]
L. Reinerman-Jones[2]
J. Lin[2]
G. Matthews[2]
D. Barber[2]
K. Dickerson[1]

U.S. Nuclear Regulatory Commission
Office of Nuclear Regulatory Research
Division of Risk Assessment
Human Factors and Reliability Branch
Washington, DC 20555-0001

Institute for Simulation and Training, University of Central Florida
3100 Technology Pkwy
Orlando, FL 32826

Niav Hughes & Kelly Dickerson, NRC Project Managers

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

HPTF RIL Series (RIL 2022-11) Preface

Much of the basis for current NRC Human Factors Engineering (HFE) guidance comes from data from research conducted in other domains (e.g., aviation, defense), qualitative data from operational experience in NPPs, and a limited amount from empirical studies in a nuclear environment. The Commission, in SRM 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 would enhance the current information gathering process, thus providing stronger technical bases and guidance to support regulatory decision making. the former Office of New Reactors (NRO) issued a user need for the Office of Nuclear Regulatory Research (RES) to update its human factors (HF) review guidance with regards to emerging technologies (User Need NRO-2012-007) and more recently the Office of Nuclear Reactor Regulation (NRR) issued a follow-on user need with the same purpose (User Need NRR-2019-008). 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 (HFRB) in the RES Division of Risk Assessment (DRA) began a program of research known as the NRC Human Performance Test Facility (HPTF) 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 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 human reliability analysis (HRA) and thus dual HF and HRA data missions.

Recognizing the essential role of data to our HF and HRA programs, the NRC historically approached data collection through multiple avenues – all with their inherent strengths and weaknesses:

1. Licensed Operators – controlled experiments at the Halden Reactor Project
2. Licensed Operators – the Scenario Authoring, Characterization, and Debriefing Application (SACADA) database capturing training scenarios
3. Novice populations – scientific literature, laboratory settings – non-nuclear

The HPTF program captures data from both novice and operational populations and the work is specifically targeted to the nuclear domain. In addition, the HPTF methodology expands upon these data collection methods. Most notably, though the addition of a new population category, that of formerly licensed operators and other nuclear domain experts. The HPTF methodology (described in detail in RIL 2022-11 Volume 1) enables the NRC to fill in the gaps from the other 3 data collection activities and conduct responsive research to support the informational needs of our users (e.g., NRR HFE technical reviewers and HRA analysts).

The intent of the HPTF was to design experiments that balanced domain realism and laboratory control sufficiently to collect systematic, meaningful, human performance data related to execution of common nuclear main control room (MCR) tasks. Three large-scale experiments were conducted to address challenges associated with developing a research methodology for using novices in a highly complex, expert driven domain. These three experiments are reported as Studies 1 and 2 in RIL 2022-11 Volume 1 which describes the approach and methodology underlying this research effort and the resulting findings for the series of studies. In RIL 2022-11 Volume 2, the Volume 1 findings were further validated via a fourth data collection by testing a formerly licensed operator population using a full-scale, full-scope simulator. Cross-experiment comparisons were enabled by leveraging a formerly licensed operator as a member of the research team to serve as senior reactor operator (SRO) and ensure participants received an experience as similar and structured as possible to the studies in Volume 1[1].

To ensure the developed methodology continues to support the HFE technical staff in user offices, the HPTF team works with those stakeholders to establish research questions and experimental design options for follow-on work. The experimental design and research questions that were examined were determined through a collaborative effort between NRC staff and a contractor with an identical simulator and performance assessment capabilities.

Toward this end, to date, three experimental design workshops have been held. The first workshop was held on March 5 and 6, 2018 upon completion of the first three HPTF experiments. The direction resulting from this first workshop was to validate the methodology and generalize the findings from the baseline HPTF experiments by using formerly licensed operators as participants to complete an experimental scenario using an analog, full-scope, full-scale simulator and a digital, part-task simulator. RIL 2022-11 Volume 2 describes the research approach and findings for the fourth experiment in the series.

The second workshop was held on August 20 and 21, 2019. The direction resulting from this second workshop was to perform a reanalysis of all HPTF experiments thus far to investigate: 1) Workload Measure Sensitivities 2) Task Order Effects and 3) Touchscreen Ergonomics. The results of each of these supplementary analyses and their regulatory implications are discussed in RIL-2022-11 Volumes 3-5 (in press). Due to the COVID-19 health crisis, the third workshop was held as a virtual series consisting of six 2-hour blocks between October 29 to November 20, 2020. The future direction topics discussed during the most recent workshop are described in RIL 2022-11 Volume 6 (in press). The final direction and experimental design are yet to be set, but the resulting methodology and results may be published as Volume 7.

These volumes of research illustrate the NRC's ongoing effort to perform systematic human performance data collection using a simulator to better inform NRC guidance and technical bases in response to Staff Requirements Memorandum (SRM) SECY-08-0195 and SRM-M061020. The HF and HRA data are essential to ensure that our HFE guidance documents and HRA methods support the review and evaluation of "state-of-the-art" HF programs (as required by 10 Code of Federal Regulations (CFR) 50.34(f)(2)(iii)).

[1] Systematic experimentation is challenging in the nuclear domain using real operators and full, dynamic scenarios because operators can take many paths to achieving a successful outcome. This variability represents a condition that is not conducive to controlled laboratory study. By including a confederate SRO in the study using a dynamic scenario, this hard to control variability is managed, thereby, enabling stable observations. See RIL 2022-11 Volumes 1 and 2 for examples of these methodological benefits.

ABSTRACT

The staff of the U.S. Nuclear Regulatory Commission (NRC) is responsible for reviewing and determining the acceptability of new reactor designs and modifications to operating plants to ensure they support safe plant operations. Human factors (HF) staff use Chapter 18 of the Standard Review plan (NUREG-0800) and the guidance documents referenced therein, in part, to ensure that plant operators can safely control the plant. The NRC's Human Factors Engineering (HFE) Program Review Model, NUREG-0711, Rev. 3 (NRC, 2012) is one of these documents. NUREG-0711 outlines that a generic "human centered" HFE design goal should include a design that supports personnel in maintaining vigilance over plant operations and provide acceptable workload levels. Furthermore, NUREG-0711's review elements highlight the importance of considering workload (WL). In particular, for Elements 3 (Task Analysis) and 4 (Staffing and Qualifications), providing an estimate of WL is explicitly part of the review criteria that must be met.

The basis for current NRC HFE guidance, in part, comes from data from research conducted in other domains (e.g., aviation, defense), qualitative data from operational experience in nuclear power plants (NPPs), and a limited amount from empirical studies in a nuclear environment. When it comes to new designs, technologies, and concepts of operations for new control rooms, there is a lack of operational experience and appropriate research literature to draw from to inform NRC HFE guidance. To address this issue, the Commission, in Staff Requirements Memorandum (SRM) SECY-08-0195, directed the staff to consider using generic simulator platforms to address human performance issues. In response to the SRM, the Office of Nuclear Regulatory Research (RES) developed the NRC Human Performance Test Facility (HPTF) research program to empirically examine human performance aspects of control room operations using a lightweight NPP simulator and a combination of objective and subjective measures of workload. The information gained will be utilized to enhance the technical basis for the NRC's regulatory guidance in HFE and to better inform models for human reliability analysis (HRA).

To date, four large-scale data collection activities have been performed for the HPTF research program. The baseline experiments are reported in RIL 2022-11 Volume 1. The purpose of this report is to present the findings of Experiment 4, which aimed to validate the methodology and generalize the findings from the baseline HPTF experiments by using formerly licensed operators as participants to complete an experimental scenario based on a generic emergency operating procedure (EOP) using an analog, full-scope, full-scale simulator and a digital, part-task simulator.

The experimental tasks were characterized into three task types: checking, detection, and response implementation. A multivariate assessment strategy using subjective and psychophysiological measures to profile and compare the workload types and levels was adopted to examine the effect of simulator type and task type. Subjective measures included the NASA-Task Load Index (NASA-TLX), Multiple Resource Questionnaire (MRQ), and Instantaneous Self-Assessment (ISA). Electroencephalogram (EEG), function Near-Infrared (fNIRS), transcranial Doppler (TCD), and Electrocardiogram (ECG) were used as the psychophysiological measures. Operators were randomly assigned in the reactor operator (RO) role or the balance of plant (BOP) role and paired into Reactor Operator teams with the Senior Reactor Operator (SRO) role filled by an experimenter. Results supported the feasibility to use digital simulators to conduct research, identify safety concerns, and may supplement operator training. The results from research using digital simulators can be conditionally generalized to

analog systems. The study also validated the methodology of using multivariate assessment to profile operator’s workload in the nuclear domain.

TABLE OF CONTENTS

LIST OF FIGURES

EXECUTIVE SUMMARY

The staff of the U.S. Nuclear Regulatory Commission (NRC) is responsible for reviewing and determining the acceptability of new reactor designs to ensure they support safe plant operations (10 CFR 50.34 (f)(2)(iii)). Human performance is a key component in the safe operation of Nuclear Power Plants (NPPs) (NRC, 2002). The human operator is a vital part of plant safety; thus, the NRC staff must understand the potential impact of new designs on human performance to make sound regulatory decisions. Much of the basis for current NRC Human Factors Engineering (HFE) guidance comes from research conducted in other domains (e.g., aviation, defense), qualitative data from operational experience in NPPs, and a limited number of empirical studies in a nuclear environment. For new designs, technologies, and concepts of operations, there is even less information. To address this information gap, the Commission in a Staff Requirements Memorandum (SRM) SECY-08-0195 directed the staff to consider using generic simulator platforms for addressing human performance issues. A simulator provides a means to gather empirical nuclear-specific human performance data that is targeted to enhancing the current information gathering process and providing stronger technical bases and guidance to support regulatory decision making. Additionally, the empirical human performance data collection ensures a better understanding of the various cognitive and physical elements that support safe control room operation.

The simulator used to address the information gap digitally represents analog instrumentation and controls (I&C) for a generic Westinghouse 3-Loop Pressurized Water Reactor controls (developed by GSE Power Systems). Using this simulator, the Human Factors and Reliability Branch (HFRB) in the Office of Nuclear Regulatory Research (RES) launched a program of experimental research with the help of the Human Performance Test Facility (HPTF) to collect empirical human performance data for measuring and understanding the various cognitive and physical elements that support safe control room operation. The intent was to design experiments that balanced domain realism and laboratory control sufficiently to collect systematic meaningful human performance data related to execution of common main control room (MCR) tasks. Investigators identified and defined three types of tasks central to the MCR: Checking, Detection, and Response Implementation. A variety of subjective and physiological measures were collected to understand the performance of those tasks in terms of both physiological and subjective workload.

The findings from the resulting experiments are presented in a series of volumes, Research Information Letter (RIL) report, “Human Performance Test Facility (HPTF) (RIL 2022-11). Volume 1, titled “Systematic Human Performance Data Collection Using Nuclear Power Plant Simulator: A Methodology” contains two studies and compares performance, physiological, and subjective measures of workload in operators and novices in a simulated digital representation of an analog plant in both a touchscreen and desktop configuration. The present report, Volume 2, titled “Comparing Operator Workload and Performance Between Digitized and Analog Simulated Environments” contains a single study and compares formerly licensed operators’ performance, physiological, and subjective workload between a full scale, full scope simulator, and the HPTF’s lightweight digitized simulator environment during an emergency operating procedure scenario.

Chapter 1 of this report begins with a description of different types of NPP simulated environments and outlines the aims of the study to validate the findings reported in RIL 2022-11

Volume 1, “Systematic Human Performance Data Collection Using Nuclear Power Plant Simulator: A Methodology” in a full-scope, full-scale simulator with hard panels and analog I&C through an examination of the effect of simulator type, operator role, and task type on operator workload. Chapter 2 provides the methodological approach and methods employed. Chapter 3 reports the results in terms of subjective, physiological, and communication performance measures. Chapter 4 contains the discussion and conclusions including the feasibility of using digital simulators to conduct research and demonstrate trends in operator workload and performance between digitized and analog simulated environments. The results of the study reported in volume 2 support that research using digital simulators can be conditionally generalized to analog systems. The study also validated the methodology of using multivariate assessments to profile operator workload in the nuclear domain due to the complex nature of NPP operations. For research purposes, multivariate assessment of workload can provide a more detailed picture of operator response to the task manipulations. Near-term future work should perform additional analyses on the data previously collected in the HPTF experimental series. The analyses resulting from this ‘deeper dive’ into the data will be published in subsequent volumes to this Research Information Letter series and will further examine:

- sensitivity of workload measures across the studies and the tasks (RIL 2022-11, volume 3)
- order effects of the tasks (RIL 2022-11, volume 4)
- potential ergonomic issues related to shoulder height and sensitivity of the touchscreens versus desktop interface control techniques (RIL 2022-11, volume 5).

ABBREVIATIONS AND ACRONYMS

ANOVA	Analysis of Variance
BWR	Boiling Water Reactor
CBFV	Cerebral Blood Flow Velocity
CFR	Code of Federal Regulation
ECG	Electrocardiogram
EEG	Electroencephalogram
EOP	Emergency Operating Procedure
fNIR	Functional Near-Infrared Spectroscopy
GPWR	Generic Pressurized Water Reactor
HFE	Human Factors Engineering
HFRB	Human Factors and Reliability Branch
HPTF	Human Performance Test Facility
HR	Heart Rate
HRA	Human Reliability Analysis
HRV	Heart Rate Variability
HSI	Human System Interface
IBI	Inter-beat Interval
I&C	Instrumentation and Control
ISA	Instantaneous Self Assessment
ISO	International Standards Organization
M	Mean
MCR	Main Control Room
Mdn	Median
MRQ	Multiple Resource Questionnaire
NPP	Nuclear Power Plant
NRC	U.S. Nuclear Regulatory Commission
OP	Operating Procedure
PSF	Performance Shaping Factor
PWR	Pressurized Water Reactor
RES	Office of Nuclear Regulatory Research
RO	Reactor Operator
SME	Subject Matter Expert
SRO	Senior Reactor Operator
TCD	Transcranial Doppler
TLX	Task Load Index
TTC	Technical Training Center

INTRODUCTION

The human performance test facility (HPTF) project aims to assess the impact of novel designs, technologies, and concept of operations on operator performance using human-in-the-loop experiments. Three experiments were conducted in HPTF phase one. Details for these experiments can be found in previous technical reports. A brief summary of these experiments is provided below.

In Experiment 1, three task types (checking, detection, response implementation) were defined for performance-based assessment. Each task type consisted of four steps that were executed using three-way communication led by the experimenter acting as the senior reactor operator (SRO). Novice participants (college students) performed common NPP operator tasks in a simplified simulated environment. A multivariate assessment strategy was adopted to profile the workload in NPP operations. Specifically, subjective measures (NASA Task Load Index (NASA-TLX), Instantaneous Self Assessment (ISA), Multiple Resource Questionnaire (MRQ)), Performance-based measures (multiple indices in three task types), and psychophysiological measures (electroencephalogram (EEG), transcranial doppler (TCD), functional near infrared spectroscopy (fNIRS), and electrocardiography (ECG)) were used in the experiment. The results suggested that numerous subjective and physiological measures were sensitive at indicating workload changes between the three task types. Generally, compared to all others, NASA-TLX frustration and fNIRS were more sensitive indices. Experiment 1 provided a systematic baseline experiment for understanding workload as it occurs in an NPP main control room (MCR) and determined the measures suitable for detecting workload responses. The results suggested that the measures were sensitive to the manipulations with novice users and supported the methodology of using novice participants in NPP MCR experimentation which can enable a larger sample size and reduce cost (Reinerman-Jones & Mercado, 2014).

In Experiment 2, touchscreens were introduced to the Human System Interface (HSI) as an independent variable to evaluate the feasibility of touchscreen technology in NPP operations and compare those outcomes to the previous experiment that used a mouse point and click desktop interface. Either touch or mouse point and click interface could be chosen for future soft control panel implementation and, therefore, it is important to understand workload levels and types associated with each. If workload is different for these HSIs this would be a consideration for continued assurance of safe operations in such integrated control rooms. The same methodology from Experiment 1 of using novice participants in a simplified simulated environment, task types, and multivariate assessment of workload were used in Experiment 2. The results suggest several potential human factors challenges in using a touchscreen interface, such as more required body (e.g., moving around) and figure (e.g., using fingers to directly manipulate controls) movement, manipulation difficulty for Instrumentation and Controls (I&C), and difference in physical feedback (Reinerman-Jones, Teo, & Harris, 2016).

Former NPP operators were recruited to participate in Experiment 3 to further validate the methodology established in Experiments 1 and 2 and confirm generalizability of workload response trends to the nuclear domain. The same task types and multivariate assessment of workload were adopted in the experiment described in this report. Due to the expertise and experience of the former NPP operators, the simplified simulator environment was restored to its original complexity of a full-scope NPP MCR simulator. Former operators in Experiment 3 and experienced participants in Experiment 2 showed similar trends in the performance and workload data. The similar trends may validate the feasibility of using student participants in a simplified simulated environment as a method for gaining insights into the performance of experienced operators in more complex simulations (Reinerman-Jones et al., 2018).

1.1 Types of NPP Simulated Environments

As NPP reactor technology and control room design has modernized and evolved, so too has the NPP simulator technology and capability. The upgrades in the NPP MCR may introduce new human factors challenges (Joe, Boring, & Persensky, 2012). The MCR houses all the I&C that control the reactor and the associated safety systems and as such can be defined as a vital area (see 10 CFR 73.2). MCR controls are defined in 10 CFR 50.54 as any apparatus that directly affects the reactivity or power output levels of the nuclear reactor (NRC, 1998). Licensed personnel are required to continually staff the reactor when it is in any operational mode other than refueling or shutdown. Per-shift on-site staffing of an MCR crew is dependent on the number of reactor units and control rooms at the NPP site and is defined in 10 CFR 55.54(2)(i). As such, there are multiple configurations and levels of capability that warrant description. In this series of HPTF experiments, the types of NPP simulated environments are characterized with the following five main features:

1.1.1 Scope

a. Full Scope Simulator – has the capability to simulate all of the physical and underlying thermodynamics occurring in the would-be plant.
b. Part Task Simulator – has the capability to simulate only part (i.e., one or two systems or just one type of task on a particular instrument or control) of plant behavior.

1.1.2 Layout

Two types of layouts were operationally defined in volume 1 of RIL 2022-11:

a. Spatially Dedicated; continuously visible – in this layout arrangement of all the Instrumentation and Controls (I&C) are available to the operator and presented in a fixed location (e.g., existing fleet of Light Water Reactors, Westinghouse pressurized water reactor (PWR)) (i.e., a one-to-one representation of the system used in real operations)

   Hierarchical – all I&C are available, but not all are continuously in view; rather they are nested and can be displayed in a hierarchical manner embedded within the workstation displays. The digital simulator depicted in figure 3 (the GSE generic pressurized water reactor, [GPWR]) has a hierarchal design.

1.1.3 Interface Type

a. Analog – conventional hard panels or bench boards with hard wired analog I&C (e.g., current training simulators for operating plants).
b. Digital – computer-based workstations with digital I&C (Westinghouse AP-1000).
c. Hybrid
   i. Digital Analog – digital representation of emulating analog I&C hard panels (e.g., GSE GPWR, the digital simulator used in the HPTF experiments).
   ii. Analog hard panels and computer-based workstations (e.g., current day control room modernization projects).

1.1.4 Workstation Design

a. Sit-down workstations
b. Stand-up workstations

1.1.5 Control Interaction Techniques

a. Mouse click input (for digital and hybrid interfaces)
b. Touch screen input (for digital and hybrid interfaces)
c. Manual manipulations of hard-wired controls (for conventional analog interfaces)

Although most operational NPPs in the United States are primarily using analog I&C in their MCR, obsolete parts are being modernized and replaced with digital interfaces, and fully digital MCRs may come online in the next few years (Joe & Boring, 2017; Hugo & Slay, 2017; Harris, Reinerman-Jones, & Teo, 2017). Currently, each operational MCR has a full-scale simulator mockup in an identical configuration (Joe & Boring, 2017). MCR designs leverage the principles in the International Standards Organization's (ISO) ergonomic design of control centers (ISO 11064-4). ISO 11064-4 specifies recommendations to follow in the ergonomic design of workstations in domains that focus on process control and security. NUREG-0700, in sections 11 and 12 includes specific details regarding a proper workstation and control room configuration. As interface technologies advance, digital simulators with full-scope thermodynamic capability become available for the industry and researchers (Hughes et al., 2017). Digital simulators with enhanced versatility and flexibility can be a potential tool for multiple purposes, such as operator training, experimentation, assessment of operator competence, human system interface evaluation, and usability tests.

In terms of above definitions, the two types of simulators used in this experiment can be characterized as the following: an analog, full-scope, full-scale simulator with spatially dedicated layout (analog full-scope/scale simulator); hybrid digital analog part-task simulator with hierarchical layout (digital part-task simulator, the GPWR, see figure 3). Both simulators require operators to remain in stand-up position during operation. The analog, full-scope/scale simulator is equipped with manual manipulations of hardwired controls, and the digital, part-task simulator utilizes touch screen input.

1.2 Validation of Findings from HPTF 1

The long-term, overarching objective for the HPTF experiment series is to examine challenges related to the impact of technology upgrades, automation of tasks, and digital interfaces on human operators' ability to perform monitoring and control functions in the MCR. In an attempt to validate the NPP simulator and methodology used in the baseline HPTF experiments and generalize the findings from the full-scope, reduced size (i.e., digital analog) simulator with a hierarchical I&C layout to the full-scope, full-size, analog hard panel simulator with a spatially-dedicated and continuously-visible I&C layout, it will be necessary to conduct a series of experiments to systematically compare the results obtained from HPTF 1, which used novices and former operators operating in a hybrid interface, to a full-scope/scale simulator consisting of traditional bench boards with hard-wired I&C.

The next logical step in generalizing the results from HPTF 1 might be to validate in a full-scale digital simulator, such as the test facility at Idaho National Lab. However, given that 1) we have readily available access to the simulators at the NRC Technical Training Center (TTC), and 2) our results from the first series of experiments are relatively consistent for two different interfaces and across the novice and more expert populations (Harris, 2017), we propose that the next step towards validation is using a full-scope, full-scale simulator with hard panels and analog I&C. While there is always value in additional data points, we contend that the saved resources warrant skipping a study using a full-scale digital simulator unless a discrepancy is observed between the results in HPTF 1 and HPTF 2, or different research questions arise in the future that would warrant doing so at that time.

The TTC Westinghouse PWR simulator is a full-scope, full-scale simulator with hard panel bench boards and an analog I&C interface. While it is used for training NRC staff on NPP technology and plant operation, it is representative of training simulators used by current operating plants. By designing a study to be conducted with the TTC simulator using a former operator participant pool but employing the same methodology in HPTF 1, we will explore the generalizability of the results from the previous studies in an environment closer to the reality of true NPP operations. Namely, while a single workload measure may provide reasonable assurance that sufficient operator resources are available to meet task demands multiple workload measures may be needed to fully characterize the many facets of operator workload in a dynamic and complex environment, like the main control room (Reinerman-Jones et al., 2018). Additionally, there was consistency across the findings from the first series of experiments. From Experiment 1 to 2 to 3, even when the population and interface change, the operational environment is staying constant and the same measures are sensitive to both students and professionals (Reinerman-Jones & Mercado, 2014; Reinerman-Jones, Teo, & Harris, 2016; Reinerman-Jones et al., 2018).

As the purpose of HPTF 1 was, in part, initially to establish a baseline, simplification was needed to be able to measure with certainty. For example, experimental control in the form of task blocking was used such that direct measurement could be done for each task type. So now, our aim is to use similar measures but, in an environment, truer to real NPP operations. More specifically, participants will perform the same task types by stepping through the full procedure/scenario without task blocking in an analog full-scope, full-scale simulator. Our methodology may be further validated if we observe similar trends for the measures and objective performance results (e.g., detection was the most demanding task type) but in a more realistic operating environment, both in terms of the simulator interface and also because they will be performing a full scenario – no blocking of task types. Furthermore, we may determine if the measures of workload are similarly sensitive to the task types and the operating environment in the same way (e.g., NASA TLX and fNIRS were the most sensitive overall regardless of task type in the HPTF 1 experiment; EEG and fNIRS were most sensitive to the detection task as a marker of the need for sustained attention).

This study would examine the same task types and use the same measures, but in a more realistic NPP environment more similar to training versus the controlled laboratory environment. Given that our results were largely consistent with two digital analog interfaces and across the samples from three populations, and we will be measuring performance on the same task types, we expect the results from HPTF 1 to view similar trends in the hard analog I&C environment. If the same trends of performance and workload data are observed, then we can be more confident in our methodology and further support our aim of generalizing to more realistic NPP operations. However, if the results we find on the three task types and workload levels from using hard analog I&C are significantly different from our existing results from HPTF 1, we will need to take a step back to use a full-scale digital simulator to figure out if it is a methodology issue or an issue with differences in the simulated environment itself.

1.3 Aim of Study

Generally, this study aimed to generalize the findings from HPTF 1 using digital analog simulators to analog systems which are still largely used in most operational NPPs in the United States for the purposes of confirming and extending the utility of the simulation for training and identifying safety concerns. More specifically, the goal for this experiment was to use a multivariate assessment strategy to profile and compare the workload types and levels in different task types across two types of simulators.

Aim 1. Examine the Effect of Simulator Type and Operator Role on Subjective Workload
Since the two simulators were of different plants, the operating procedures were modified to be comparable, but not identical. Although the digital simulator offered a digital representation of emulating analog I&C hard panels, the extra steps and additional panels involved in the analog, full-scope/scale simulator condition may elicit higher workload. The goal of aim 1 was to validate previous workload findings from GPWR studies on a more operationally realistic simulator. It was therefore, hypothesized that participants in the analog, full-scope/scale simulator condition and in the role of BOP should report higher subjective workload in the subscales linked to temporal demand and spatial process.

Aim 2. *Examine the Effect of Simulator Type and Task Type on Physiological Indices of Workload*
Multiple physiological measures, including EEG, ECG, TCD, and fNIRS, were applied in this study. Generally, it was hypothesized that participants in the analog, full-scope/scale simulator condition should show similar trends in terms of task type in the physiological indices as the participants in the digital, part-task simulator condition.

Aim 3. Examine the Effect of Simulator Type and Task Type on Communication in Operation
Four communication metrics, including percent communications completed correctly, number of I&C location help requests, number of clarifications required, and number of requests for repeating an instruction were recorded as the performance measures. It was hypothesized that participants in the analog, full-scope/scale simulator condition should have more errors and request for support due to the additional steps in operational procedure and the more panels involved in operation.

2 METHODOLOGY

2.1 Participants

A total of 48 participants (24 crews, *M* = 51.90, *SD* = 10.02) were recruited through an agency announcement that was distributed to NRC employees and trainees at the TTC. Nine crews (14 males, 4 females, *M* = 45.94, *SD* = 10.64) participated in the study using the digital, part-task simulator; fifteen crews (30 males, *M* = 55.47, *SD* = 7.82) participated in the study in the analog, full-scope/scale simulator. All participants for this study were former NPP operators who had operational experience working in an MCR from the commercial nuclear power generation and/or naval nuclear power generation domains. These participants consisted of former PWR and/or boiling water reactor (BWR) operators. All participants reported normal health status and no history of neurological disorders. Participants were compensated by their regular hourly wage.

2.2 Training

This training involved a demonstration using PowerPoint slides and a hands-on training of the simulator to make sure that the participants were familiar with the system for the assigned simulator condition. Additionally, participants went through a three-way communication training and practice with the SRO to clearly relay critical information, navigate within the simulator to locate and read status indicators, and respond appropriately to a simulated NPP system warning by following procedures. Because operational procedures may vary among different NPP MCRs, this training session was used to ensure all participants were familiarized with the expected procedures and configuration of the assigned simulator for the experiment.

2.3 Experimental Design

A 2 (Simulator: analog, full-scope/scale and digital, part-task) × 2 (Operator role: Reactor Operator (RO) and BOP) × 3 (Task type: checking, detection, and response implementation) mixed factorial design was adopted in this study. The simulator and the operator role were between-groups factors. The task type was a within-subjects factor. A team of two randomly assigned participants was paired for each experimental session. One participant performed the duties assigned to the role of RO and the other performed the duties assigned to the role of BOP. The steps in the experiment procedure were categorized into three task types, in terms of checking, detection, and response implementation.

2.4 Independent Variable

The independent variables in this experiment were simulator type (analog, full-scope/scale and digital, part-task), task type (checking, detection, and response implementation) and operator role (RO and BOP).

2.4.1 Task Type

The task type consisted of three conditions. The checking task type required a one-time inspection of an I&C to verify that it was in the state called for by the EOP. Participants were required to locate various I&Cs and indicate identification by clicking on the correct I&C. The detection task type required participants to correctly locate an instrument and then to continuously monitor that instrument parameter for identification of change. The response implementation task type required participants to locate a control and subsequently manipulate the control in the required direction (i.e., open or shut).

2.5 Simulator

Experimental sessions for the analog, full-scope/scale simulator condition were conducted using a simulator at the TTC in Chattanooga, Tennessee, which replicates a Westinghouse 4-Loop PWR design and has the capability to simulate all the physical and underlying thermodynamics occurring in the real plant (Figure 1). The simulator was originally designed to be used to train operators in a nuclear power plant. It was taken from a decommissioned plant and repurposed to train NRC staff at the TTC. It is used at the TTC in diverse and demanding programs including: training of inspectors, operator license examiners, and other technical personnel in transient and accident response; extended scenarios including prolonged operation in Emergency Operating Procedures (EOP's); "what if" (best estimate) analyses; and support of various human factors and research projects (NUREG-1527, 1995). The simulator 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.

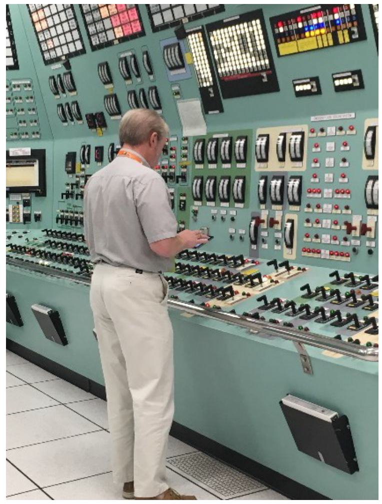

Figure 1 The analog, full-scope/scale simulator

Experimental sessions for the digital, part-task simulator condition were conducted in a laboratory room, set up as a mock MCR at the NRC headquarters in Rockville, Maryland (Figure 2). This GSE Generic PWR NPP MCR simulator consisted of eight 27-inch touchscreen monitors arranged two high by two wide. Each monitor had a resolution of 2560 by 1440 pixels.

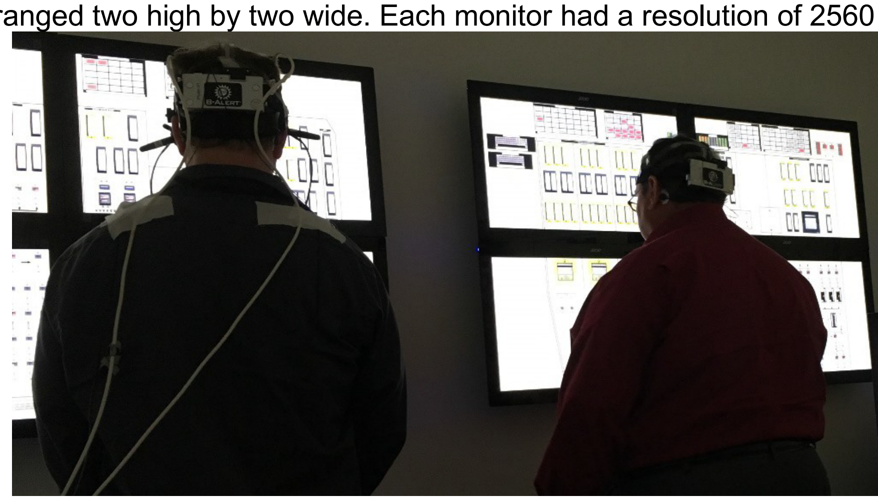

Figure 2 Operators operating on the digital, part-task simulator.

Both simulators were configured for a crew of three operators. Crews consisted of an SRO, an RO and a BOP, whereby participants operated in the role of either RO or BOP and a researcher performed the role of SRO.

2.6 <u>Physiological Instruments</u>

A suite of three physiological instruments (Advanced Brain Monitoring's B-Alert X10, Spencer Technologies' ST3 Digital Transcranial Doppler, and Somantics' Invos Cerebral/Somatic Oximeter) were used to monitor workload states during experimental sessions for the RO. For the BOP, workload state was monitored using the Advanced Brain Monitoring's B-Alert X10.

Details about the specific signals these systems monitored are explained in the physiological measures section below.

2.7 Scenario Setup

The experimental scenario was developed based on a generic version for a “Loss of All Alternating Current (AC) Power (ECA-0.0)” EOP but modified for experimental use. The two simulators were of different plants. NPP MCR designs have been developed and modified over many years; as a result, most NPP MCR I&C layouts and workstation configurations are unique. Therefore, the operating procedures were modified to be comparable, but not identical.

2.7.1 Analog, full-scope full-scale simulator type condition

The experiment procedure contained 69 steps supporting three different task types, i.e., checking, detection, and response implementation task types. In the experiment procedure, there were 30 steps (16 checking, 5 detection, and 9 response implementation) for the RO, and 39 steps (27 checking, 1 detection, and 11 response implementation) for the BOP. The number of steps were not balanced for the RO nor task type due to the nature of the original EOP, which requires steps to be taken in a prescribed sequence. In case the crew made an error or took alternative actions outside the scope of experimental procedures, an original EOP was available to the SRO for use as a contingency plan. However, the contingency plan was never required in the actual experiments.

2.7.2 Digital, part-task simulator type condition

The entire experiment procedure contained 48 steps. Both the RO and BOP were required to complete 24 steps (8 checking, 8 detection, and 8 response implementation). The order of steps was partially counterbalanced to keep the ecological validity based on the original EOP (Hughes et al., 2017).

2.8 Demographics

A demographics questionnaire was developed to gather information about age, gender, education, health state, and experience in different types of reactors.

2.9 Communication Performance

A voice recorder recorded verbal three-way communication between the operators (RO/BOP) and the SRO. Three-way communication performance measures included instruction events per task, instruction events repeated, instruction clarifications, and location help. Instruction events per task were the number of three-way communication events completed. An instruction event repeated was the number of requests by participants for a repeated instruction and the number of requests by the SRO for a repeated response from participants. An instruction clarification was a clarification by the SRO to a participant. Location help was the number of requests, by participants, for assistance in locating the correct control.

2.10 Subjective Measures

2.10.1 NASA-Task Load Index (NASA-TLX)

The NASA - Task Load Index (NASA-TLX; Hart & Staveland, 1988) is a widely used multi-dimensional measurement of subjective workload. In this study, it was used to measure the perceived workload at the end of each task type or after the entire experimental scenario, depending on the simulator condition. It consists of six rating scales for workload-relevant factors, including mental demand, physical demand, temporal demand, performance, effort, and frustration. All factors, except performance, are rated on a 0 - 100 scale from “Low” to “High”. Performance is rated on a 0 - 100 scale from “Good” to “Poor”.

2.10.2 Multiple Resource Questionnaire (MRQ)

The MRQ was used to characterize the nature of the mental processes used during a task (Boles & Adair, 2001). The items on the questionnaire were derived from factor analytic studies of lateralized processes (Boles, 1991, 1992, 1996, 2002). Participants received a copy of the scale with definitions and completed the MRQ at the end of each task type or the scenario depending on the simulator condition using a computerized version of the questionnaire. The following 14 of 17 scales were included for the present study: auditory emotional process, auditory linguistic process, manual process, short-term memory process, spatial attentive process, spatial categorical process, spatial concentrative process, spatial emergent process, spatial positional process, spatial quantitative process, visual lexical process, visual phonetic process, visual temporal process, and vocal process.

2.10.3 Instantaneous Self-Assessment (ISA)

The Instantaneous Self-Assessment (ISA: Tattersall & Foord, 1996) is a subjective unidimensional workload rating method that provides a continuous and concurrent assessment of task demand on perceived workload. In this study, it was used to measure the perceived workload for each task type. Participants were asked to verbally rate their workload for completing each of the three task types (checking, detection, and response implementation) using a 5-point Likert scale ranging from “1 = Very Low” to “5 = Very High”.

2.11 Physiological Measures

2.11.1 Electroencephalogram (EEG)

The Advanced Brain Monitoring B-Alert X10 system was employed to assess nine-channels of EEG and one channel of ECG (Figure 3). Following the international standard 10-20 System, the sampling rate of 256 Hz captures signals from Fz, F3, F4, Cz, C3, C4, Pz, P3, and P4. Reference electrodes were placed on each participant’s mastoid bone. PSD analysis techniques was used to analyze three standard bandwidths: theta (4-8 Hz), alpha (9-13 Hz), and beta (14-30 Hz; Wilson, 2002). Each bandwidth was collected for the nine nodes. They are combined to compare left and right hemispheres and the front, temporal, and parietal lobes.

Figure 3 ABM's X10 EEG/ECG system

2.11.2 Electrocardiogram (ECG)

The Advanced Brain Monitoring System B-Alert X10 system was used to monitor the ECG, sampling at 256 Hz. Single-lead electrodes were placed on the center of the right clavicle and one on the lowest left rib (Figure 4). Heart Rate (HR) was computed using peak cardiac activity to measure the interval from each beat per second. “So and Chan” QRS detection methods were used to calculate Interbeat Interval (IBI) and Heart Rate Variability (HRV: Taylor, Reinerman-Jones, Cosenzo, & Nicholson, 2010). This approach maximizes the amplitude of the R-wave (Henelius, Hirvonen, Holm, Korpela, & Muller, 2009).

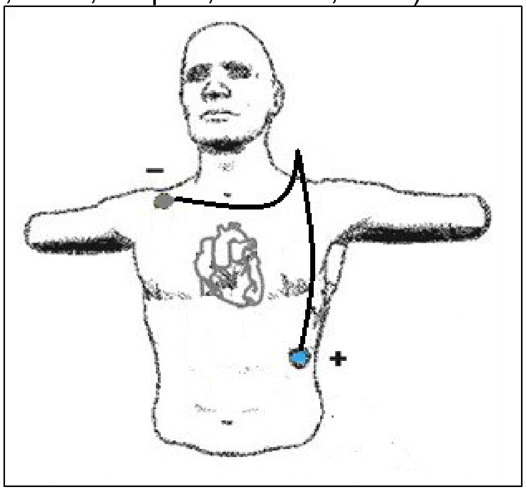

Figure 4 Electrode locations for the ECG system

2.11.3 Transcranial Doppler (TCD)

The Spencer Technologies’ ST3 Digital Transcranial Doppler, model PMD150, was used to monitor cerebral blood flow velocity (CBFV) of the medial cerebral artery (MCA) in the left and right hemisphere through high pulse repetition frequency (PRF; Figure 5). The Marc 600 head frame set was used to hold the TCD probes in place.

Figure 5 Spencer Technologies' ST3 Transcranial Doppler

2.11.4 Functional Near Infrared Spectroscopy (fNIRS)

The Somantics' Invos Cerebral/Somatic Oximeter, model 5100C (Figure 6), was used to monitor (hemodynamic) changes in oxygenated hemoglobin (oxy-Hb) and deoxygenated hemoglobin (deoxy-HB) in the left and right hemisphere prefrontal cortex (Ayaz et al., 2011; Chance, Zhuang, UnAh, Alter, & Lipton, 1993).

Figure 6 fNIRS screen

2.12 <u>Procedures</u>

Before the experiment sessions, an informed consent agreement was received by researchers. Then, the participants were asked to turn off cell phones and remove watches. Next, participants were instructed to complete the pre-task survey set, including the Demographic Questionnaire and the pre-task Dundee Stress State Questionnaire (DSSQ). The total time for pre-task activities was approximately 30 minutes.

After completing pre-task surveys, training started with a demonstration of the simulation to make sure that the participants were familiar with the system, followed by a three-way communication training and practice with the SRO. Once the participants completed the training, researchers then connected them to the physiological sensors, depending on their role. The RO was connected to fNIRS, TCD, EEG, ECG sensors and Microsoft Band II. The BOP was only connected to EEG, ECG sensors and Microsoft Band II since the wired fNIRS and TCD sensors would obstruct them from performing their tasks. Once both participants were connected to the sensors, a five-minute wakeful rest baseline was conducted simultaneously for the two paired participants. The training and physiological sensor connection took approximately 90 minutes.

After the baseline, the emergency event (loss of all AC power) was triggered by the simulator engineer. Then, the SRO started the experimental scenario and followed the experimental procedures to task the RO and BOP using three-way communication. During the experiment session, three researchers were following along with the SRO, RO, and BOP respectively to track their performance. The experiment session took approximately 60 minutes.

The ISA ratings were prompted after certain steps during the scenario. For the digital part-task simulator condition, participants were instructed to complete the post-task surveys, including NASA-TLX, post-task DSSQ, and MRQ after each task type. Whereas for the analog, full-scope/scale simulator condition, the post-task surveys were provided after the experiment scenario due to the nature of the experimental EOP.

3 RESULTS

3.1 Subjective Measures

3.1.1 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, such that participants in the digital, part-task simulator condition (M = 34.72, SD = 21.50) reported higher frustration ratings than participants in the analog, full-scope/scale simulator condition (M = 17.83, SD = 21.44), $F(1,44) = .12$, $p < .05$, , $\eta_p^2 = .14$. 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, $F(1,44) = 5.38$, $p < .05$, $\eta^2 = .11$; 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 (Figure 7).

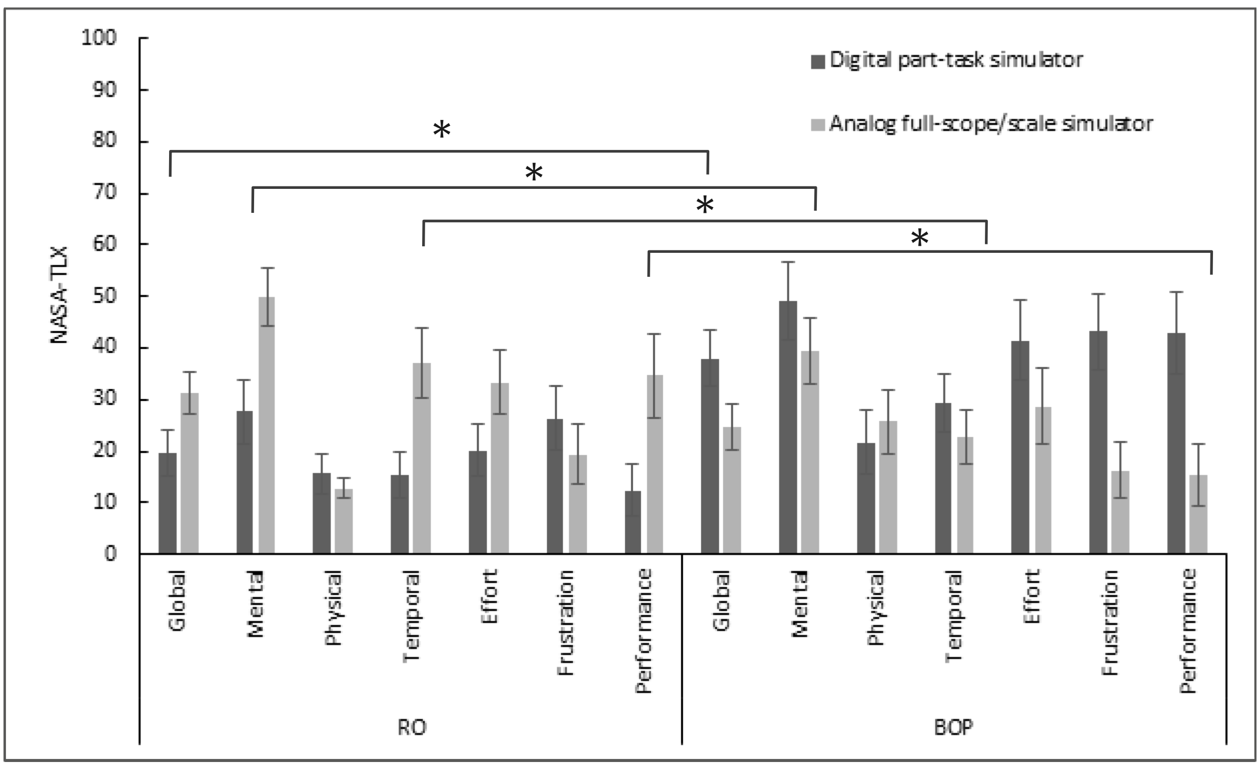

Figure 7 NASA-TLX global and subscale scores (error bars denote standard errors). Asterisks denote significant effects.

3.1.2 MRQ

A series of 2 (Simulator: digital, part-task and analog, full-scope/scale) × 2 (Operator Role: RO and BOP) ANOVAs were run for the 17 subscales in MRQ to test the effects of experimental manipulations on various mental processes that contributed to the workload experienced.

Operator role showed a significant main effect on manual process, $F(1,44) = 4.21$, $p < .05$, $\eta_p^2 = .09$ and spatial attentive process, $F(1,44) = 5.69$, $p < .05$, $\eta_p^2 = .12$. Participants in BOP role reported higher ratings in the manual process ($M = 66.07$, $SD = 21.98$) and spatial attentive process ($M = 78.75$, $SD = 14.88$) than participants in the RO role (manual: $M = 49.74$, $SD = 23.15$; spatial attentive: $M = 68.14$, $SD = 18.37$). In addition, results also revealed a main significant effect of simulator type on spatial positional process, $F(1,44) = 6.22$, $p < .05$, $\eta_p^2 = .12$; spatial quantitative process, $F(1,44) = 7.81$, $p < .01$, $\eta_p^2 = .15$; visual temporal process, $F(1,44) = 5.47$, $p < .05$, $\eta_p^2 = .11$; and vocal process, $F(1,44) = 8.58$, $p < .01$, $\eta^2 = .16$. Participants in the analog, full-scope/scale simulator condition showed greater ratings in the four mental processes than participants in the digital, part-task simulator condition (Figure 8). No interaction between simulator type and operator role was significant for any of the subscales in MRQ.

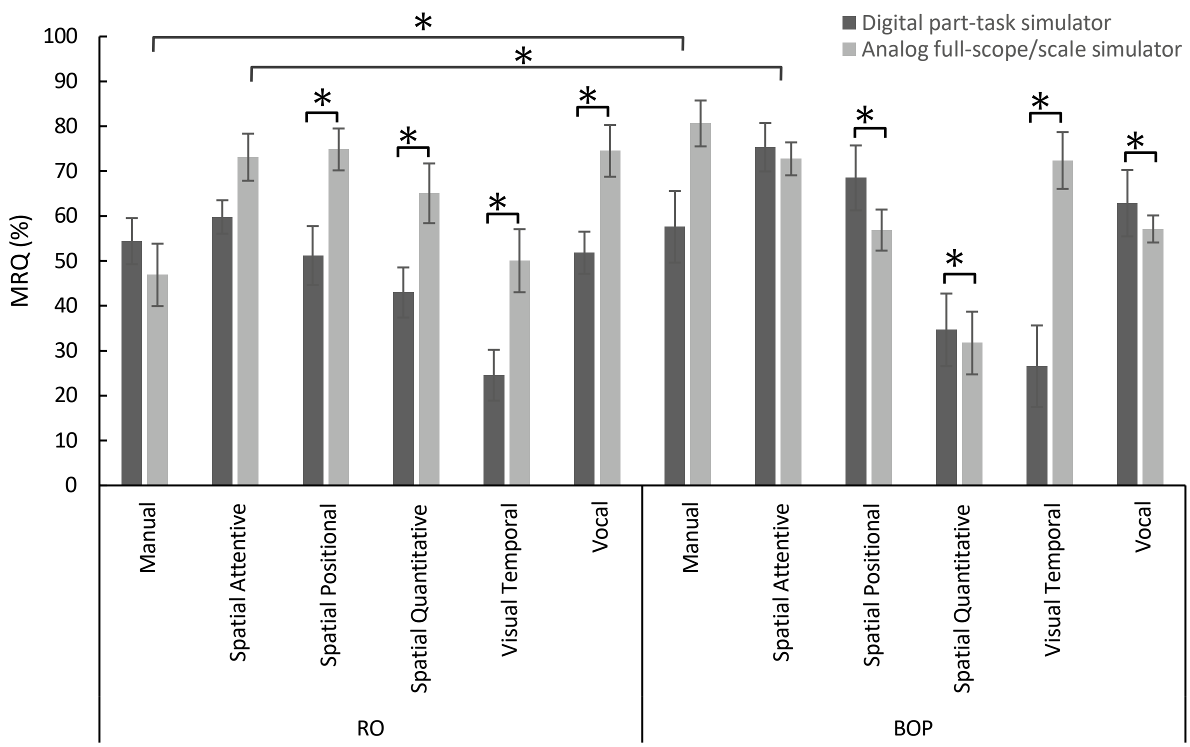

Figure 8 Significant MRQ ratings (error bars denote standard errors). Asterisks denote significant effects.

3.1.3 ISA

A 2 (Simulator: digital, part-task and analog, full-scope/scale) × 2 (Operator Role: RO and BOP) × 3 (Task Type: checking, detection, and response implementation) mixed ANOVA with repeated measures on task type was run to test the effects of experimental manipulations on the workload experienced. A significant main effect of simulator type was revealed, $F(1,44) = 15.54$, $p < .01$, $\eta_p^2 = .26$. Participants in the analog, full-scope/scale simulator condition reported higher ISA workload in all three task types than participants in the digital, part-task simulator. The main effect of task type was also significant, $F(1.76,77.56) = 4.07$, $p < .05$, $\eta_p^2 = .09$. Overall, participants rated the detection task type as the most workload-demanding task type. Additionally, the interaction of task type and operator role was significant, $F(1.76,77.56) = 8.00$, $p < .01$, $\eta_p^2 = .15$. The detection task type was rated as the highest workload task type in the digital, part-task simulator condition and in the analog, full-scope/scale simulator condition by

ROs, however, this trend was not seen in participants in the BOP role in the analog, full-scope/scale simulator condition (Figure 9).

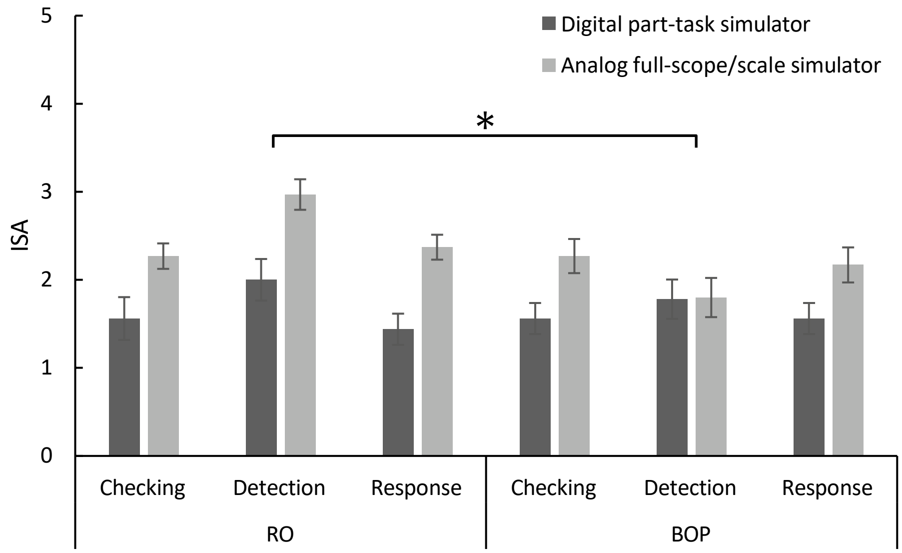

Figure 2 ISA ratings (error bars denote standard errors). Asterisk denotes significant effects.

3.2 Physiological Measures

All physiological dependent variables entered into the ANOVAs were percentages, based on difference scores, calculated from a five-minute resting baseline. For example, if the participant's left cerebral blood flow velocity (CBFV) for the five-minute baseline was 72.41 cm/s and their left CBFV for the subsequent checking task was 74.51 cm/s, their raw difference from baseline would be 2.10 cm/s, and the percentage change would be 2.90%. This method helps account for individual differences when comparing group means as is the case when running ANOVAs.

3.2.1 Physiological Measures Electroencephalogram (EEG)

Brain activity was recorded at 9 EEG sensor sites, the EEG data was analyzed by grouping sensor sites by hemispheres (i.e., compare brain activity between the left and right hemispheres) and lobes (i.e., compare brain activity among the frontal, parietal and occipital lobes).

3.2.1.1 Hemispheres

Three (Task Type: checking, detection, and response implementation) × 2 (Hemisphere: left and right) × 2 (Simulator: digital, part-task and analog, full-scope/scale) mixed ANOVAs were run for theta, alpha, and beta frequency bands separately. These ANOVAs provided insight into the overall effects of task type and simulator type on the left and right hemispheres.

For theta band, a significant main effect was found for task type, $F(2,84) = 8.35$, $p < .01$, $\eta_p^2 = .16$, such that detection (M = 60.51 SD = 100.10) elicited smaller theta increase than checking (M = 88.68 SD = 111.11), and response implementation (M = 74.66 SD = 93.89). In addition, the main effect of simulator was also significant, $F(1,42) = 4.65$, $p < .05$, $\eta_p^2 = .10$. In general, the

EEG theta waves observed from the participants in different simulator conditions showed a similar trend across task types and hemispheres, but the changes in the analog full-scope/scale simulator were in a greater magnitude (Figure 10). No significant main effect for hemisphere and no interaction was found for theta waves.

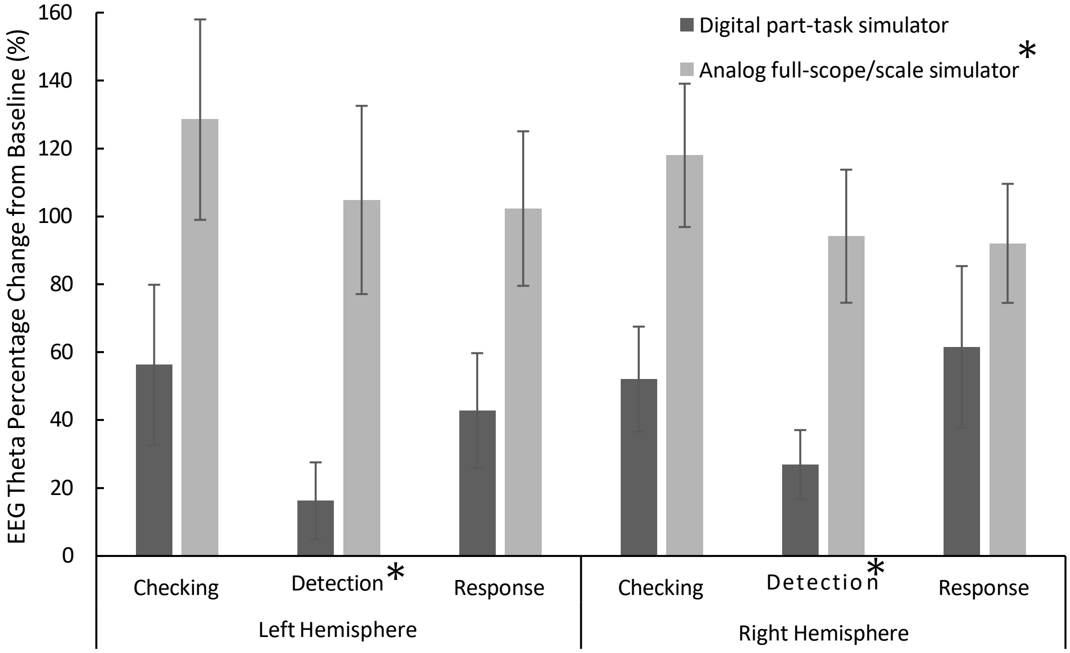

Figure 3 EEG theta percentage change from baseline by task type and hemisphere (error bars denote standard errors)

In terms of alpha band, a significant main effect was found for task type, $F(1.75,73.29) = 5.37$, $p < .01$, $\eta_p^2 = .13$. Overall, the detection task type ($M = 37.08$ $SD = 83.04$) showed a smaller increase from baseline than the checking task type ($M = 52.34$ $SD = 97.77$). The main effect of simulator was also significant, $F(1,42) = 7.19$, $p < .05$, $\eta_p^2 = .15$. In general, EEG alpha waves observed from the participants in different simulator condition showed a similar trend across task types and hemispheres, but the changes in the analog full-scope/scale simulator were in a greater magnitude (Figure 11). No significant main effect for hemisphere and no interaction was found for alpha waves.

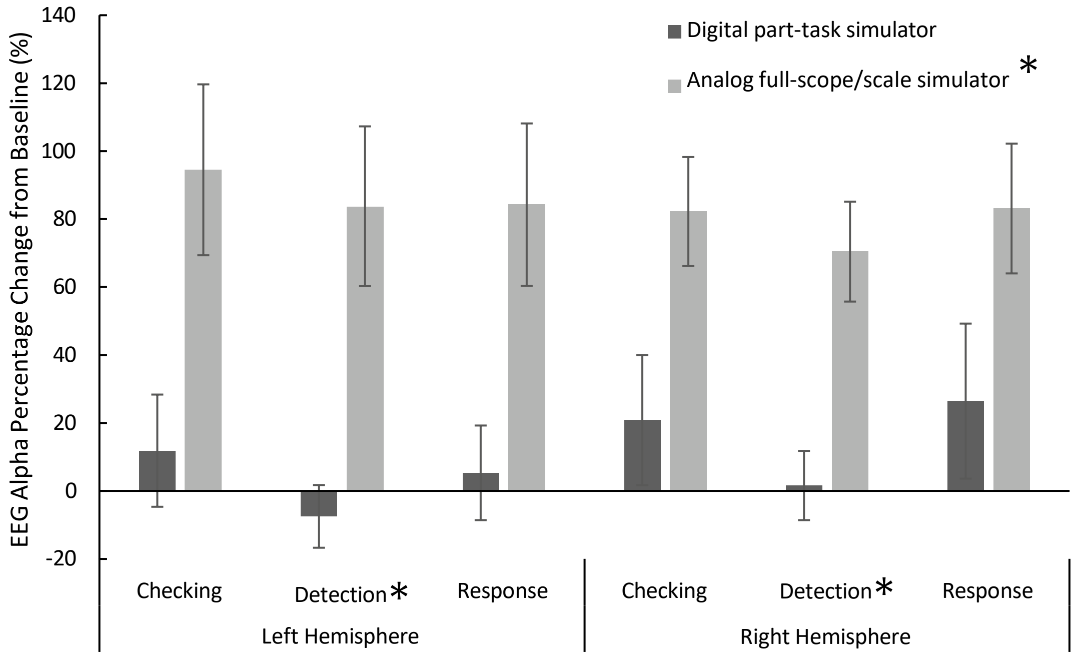

Figure 4 EEG alpha percentage change from baseline by task type and hemisphere (error bars denote standard errors)

For EEG beta band, a significant main effect was found for task type, $F(2,84) = 7.63$, $p < .01$, $\eta_p^2 = .15$. Compared to the checking task type (M = 88.71 SD = 72.60), the detection task type (M = 69.84 SD = 69.12) showed a smaller increase from baseline. The main effect of simulator type was also significant, $F(1,42) = 14.35$, $p < .01$, $\eta_p^2 = .26$. EEG beta waves observed from the participants in different simulator condition showed a similar trend across task types and hemispheres, but the changes in the analog full-scope/scale simulator were in a greater magnitude. In addition, the main effect of hemisphere was significant for EEG beta band, $F(1,84) = 16.04$, $p < .01$, $\eta_p^2 = .28$. Generally, left hemisphere beta showed a greater increase than right hemisphere beta. No significant interaction was found for beta waves (Figure 12).

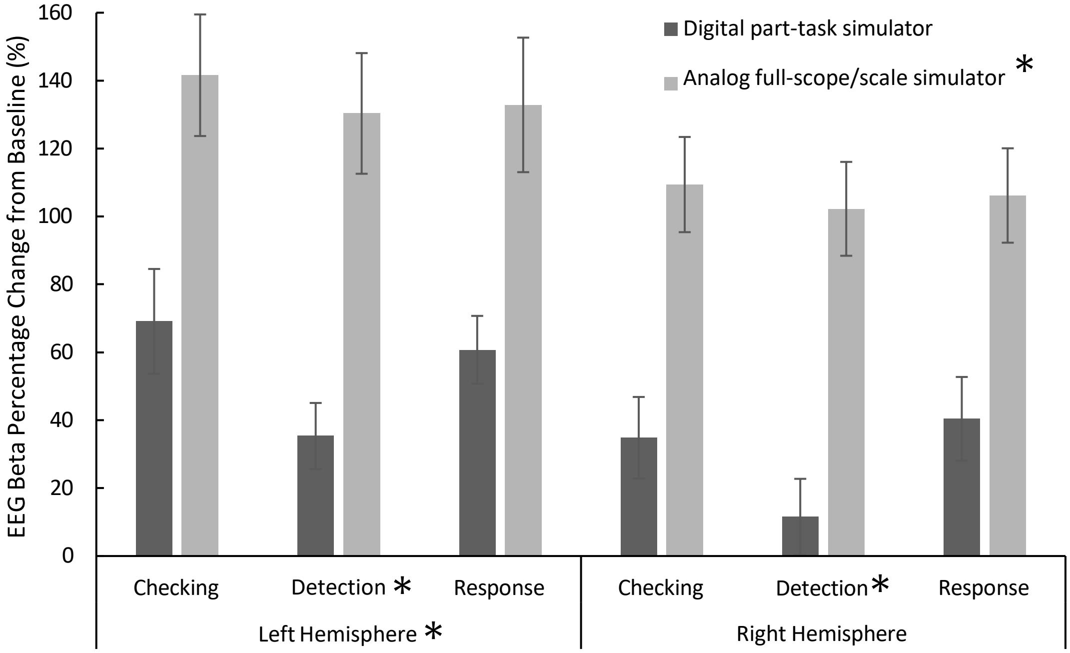

Figure 5 EEG beta percentage change from baseline by task type and hemisphere (error bars denote standard errors)

3.2.1.2 Lobes

Three (Task Type: checking, detection, and response implementation) × 3 (Lobe: frontal, parietal, and occipital lobes) × 2 (Simulator: digital, part-task and analog, full-scope/scale) mixed ANOVAs were run for EEG theta, alpha, and beta frequency bands separately. These ANOVAs provided insight into the overall effects of task type and simulator type on the frontal, parietal, and occipital lobes.

For theta frequency band, a significant main effect of task type was found, $F(2,84) = 8.07$, $p < .01$, $\eta_p^2 = .16$, such that detection task type (M = 58.38 SD = 86.36) showed a lower increase than checking task type (M = 82.48 SD = 98.66). In addition, the main effect of simulator type was significant, $F(1,42) = 4.72$, $p < .05$, $\eta_p^2 = .10$. EEG theta waves observed from the participants in different simulator conditions showed a similar trend across task types and lobes, but the changes in the analog full-scope/scale simulator were in a greater magnitude. No significant interaction was found (Figure 13).

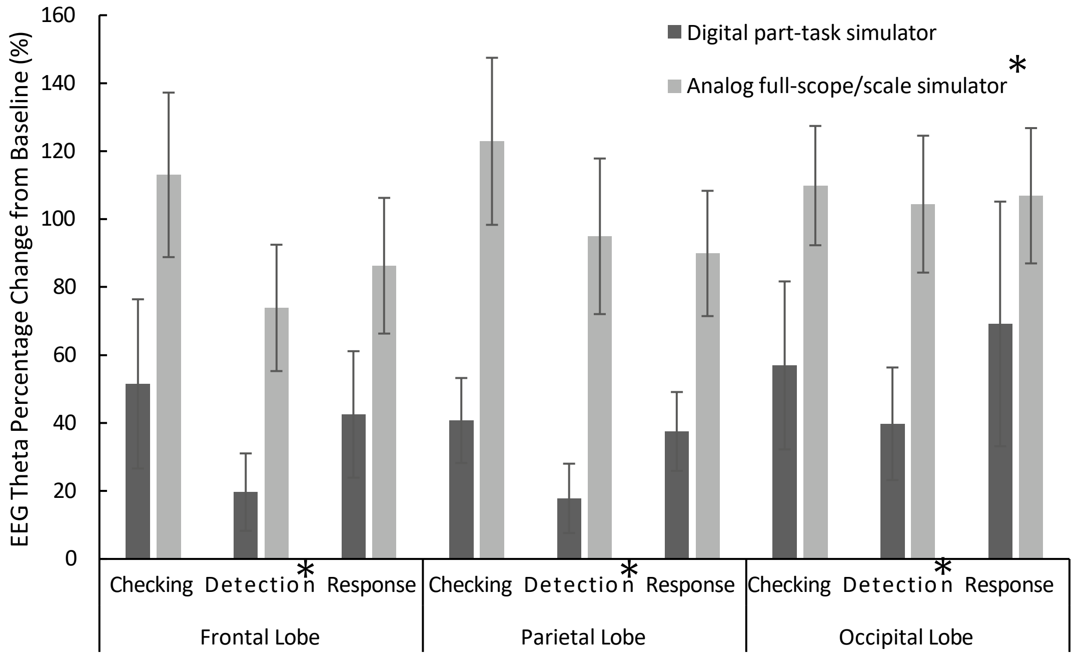

Figure 6 EEG theta percentage change from baseline by task type and lobe (error bars denote standard errors)

For alpha frequency band, a significant main effect was found for task type, $F(1.60,67.16) = 4.47$, $p < .05$, $\eta_p^2 = .10$. Overall, the detection task type ($M = 35.37$, $SD = 77.53$) showed a smaller increase from baseline than the checking task type ($M = 49.26$ $SD = 91.15$). The main effect of simulator was also significant, $F(1,42) = 7.39$, $p < .01$, $\eta_p^2 = .15$. In general, EEG alpha waves observed from the participants in different simulator conditions showed a similar trend across task types and hemispheres, but the changes in the analog full-scope/scale simulator were in a greater magnitude (Figure 14). No significant main effect for lobe and no interaction was found for alpha waves.

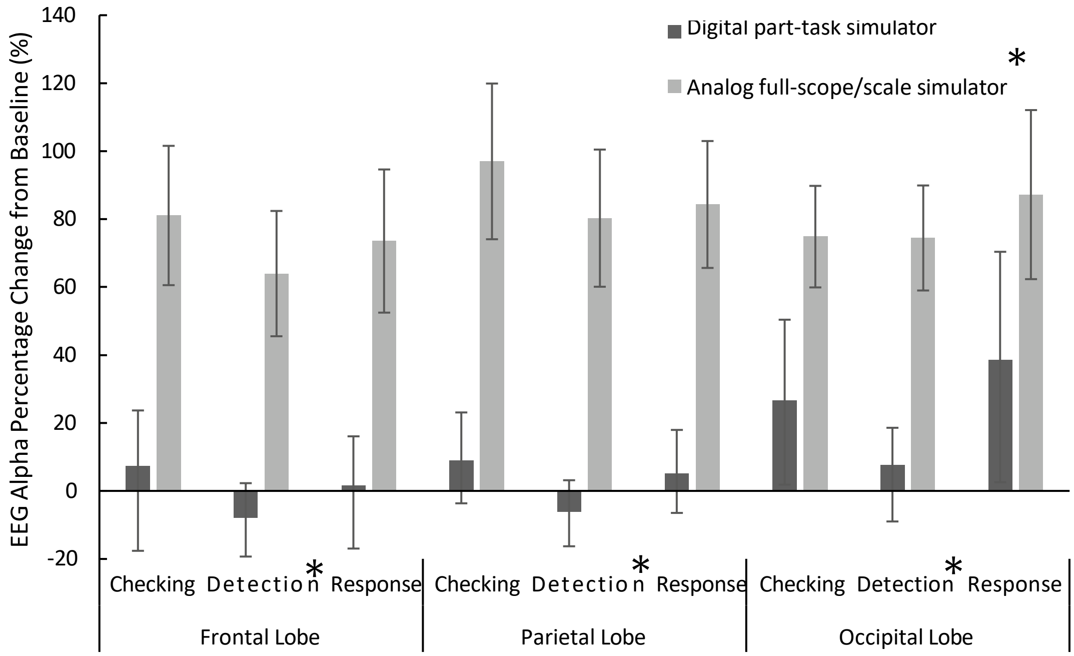

Figure 7 EEG alpha percentage change from baseline by task type and lobe (error bars denote standard errors)

For EEG beta band, a significant main effect was found for task type, $F(2,84) = 8.32$, $p < .01$, $\eta_p^2 = .17$. The detection task type (M = 68.45 SD = 68.54) showed a smaller increase from baseline than both the checking task type (M = 87.14 SD = 70.75) and the response implementation task type (M = 84.58, SD = 72.73). The main effect of simulator type was also significant, $F(1,42) = 14.50$, $p < .01$, $\eta_p^2 = .26$. EEG beta waves observed from the participants in different simulator conditions showed a similar trend across task types and lobes, but the changes in the analog full-scope/scale simulator were in a greater magnitude. In addition, the main effect of lobe was significant for EEG beta band, $F(1.42,59.66) = 4.73$, $p < .05$, $\eta_p^2 = .10$. Generally, frontal lobe beta showed a greater increase than occipital lobe beta (Figure 15). No significant interaction was found for beta waves.

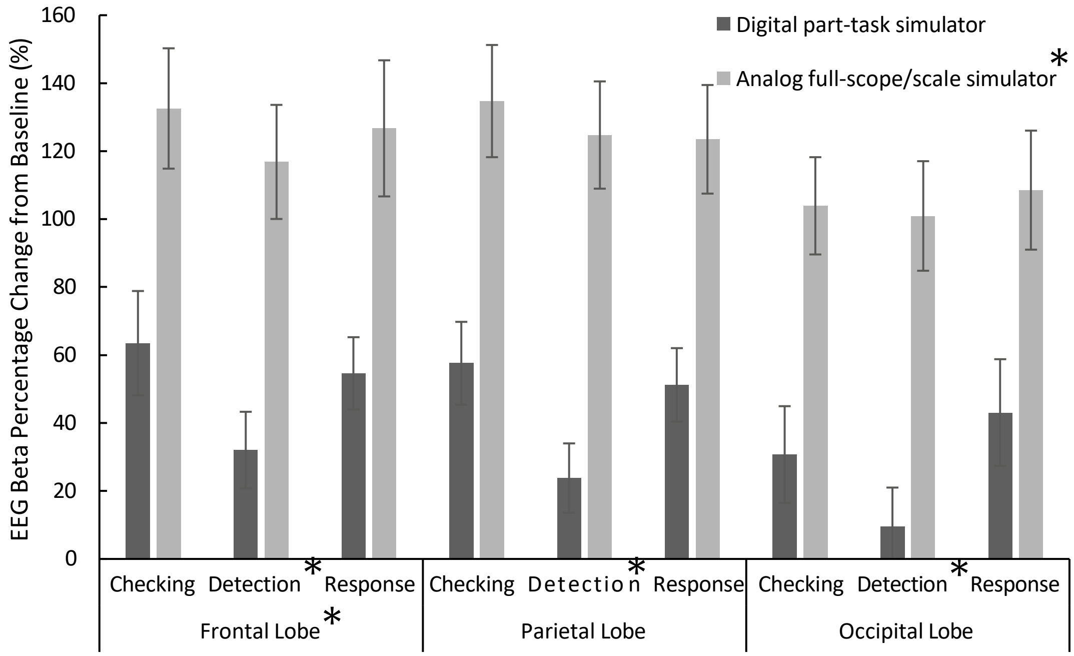

Figure 8 EEG beta percentage change from baseline by task type and lobe (error bars denote standard errors)

3.2.2 Transcranial Doppler Ultrasonography (TCD)

Due to the restriction of the sensor connection, the TCD sensor was only available for data collection on 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 mean CBFV. The ANOVA results revealed no significant main effect for task type or simulator type, and no significant interaction.

3.2.3 Functional Near-Infrared Spectroscopy (fNIRS)

Due to the restriction of the sensor connection, the fNIRS sensor was only available for data collection on 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 (rSO_2) for the left and right prefrontal cortex. A main effect for task type was found, $F(2,32) = 5.48$, $p < .01$, $\eta_p^2 = .26$. Compared to the response implementation task type ($M = 1.38$ $SD = 2.18$), the detection task type ($M = 2.22$ $SD = 2.17$) showed a greater increase from baseline (Figure 16). No simulator effect or interaction was found.

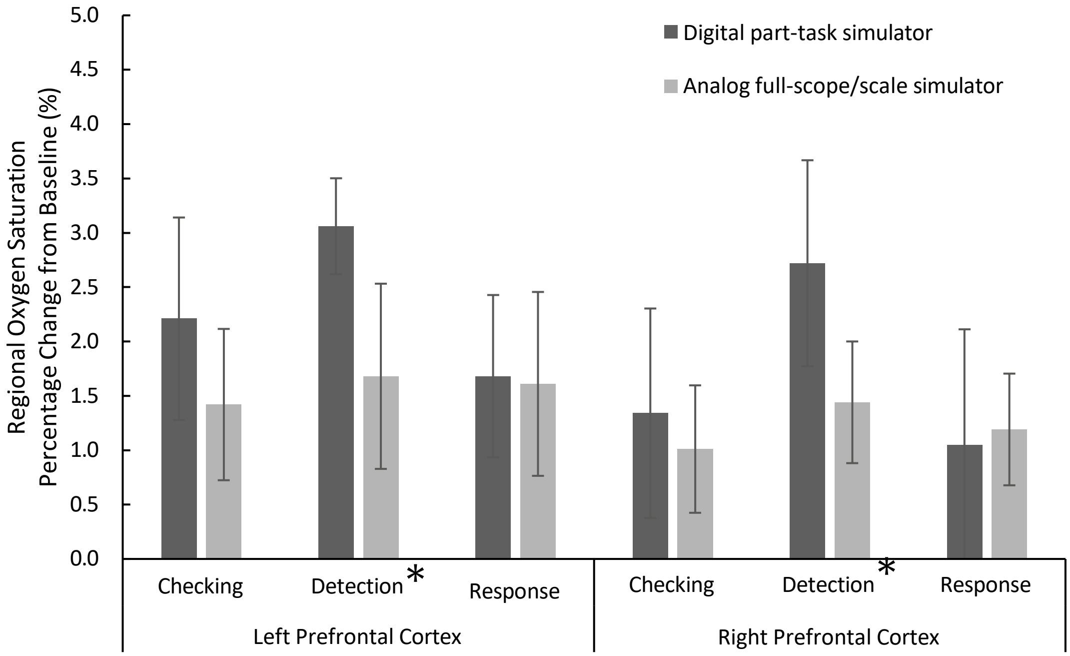

Figure 9 Pre-frontal cortex rSO2 percentage change from baseline by task type and hemisphere (error bars denote standard errors)

3.2.4 Electrocardiogram (ECG)

Three (Task Type: checking, detection, and response implementation) × 2 (Simulator: digital, part-task and analog, full-scope/scale) mixed ANOVAs with repeated measures on task type were conducted to determine the effects of task types and simulator types on HR, HRV, and IBI. The ECG metrics, HR, IBI, and HRV, were derived from R-Peak detections using the So-Chan QRS algorithm from the raw ECG signal. For the metric of HR, the interaction of task type and simulator type was significant, $F(1.55,60.26) = 6.64$, $p < .01$, $\eta_p^2 = .15$. In the digital part-task simulator condition, participants showed higher increases in heart rate during the detection tasks, whereas this trend was not observed in the analog full-scope/scale simulator condition (Figure 17). No significant main effect of task type or simulator type was found. For the metric of IBI, the interaction of task type and simulator type was significant, $F(1.61,62.82) = 7.53$, $p < .01$, $\eta_p^2 = .16$. In the digital part-task simulator condition, participants showed greater decrease in interbeat interval during the detection tasks, but this trend was not observed in the analog full-scope/scale simulator condition (Figure 18). No significant main effect of task type or simulator type was found. In addition, no significant main effect or interaction was found for the metric of heart rate variability.

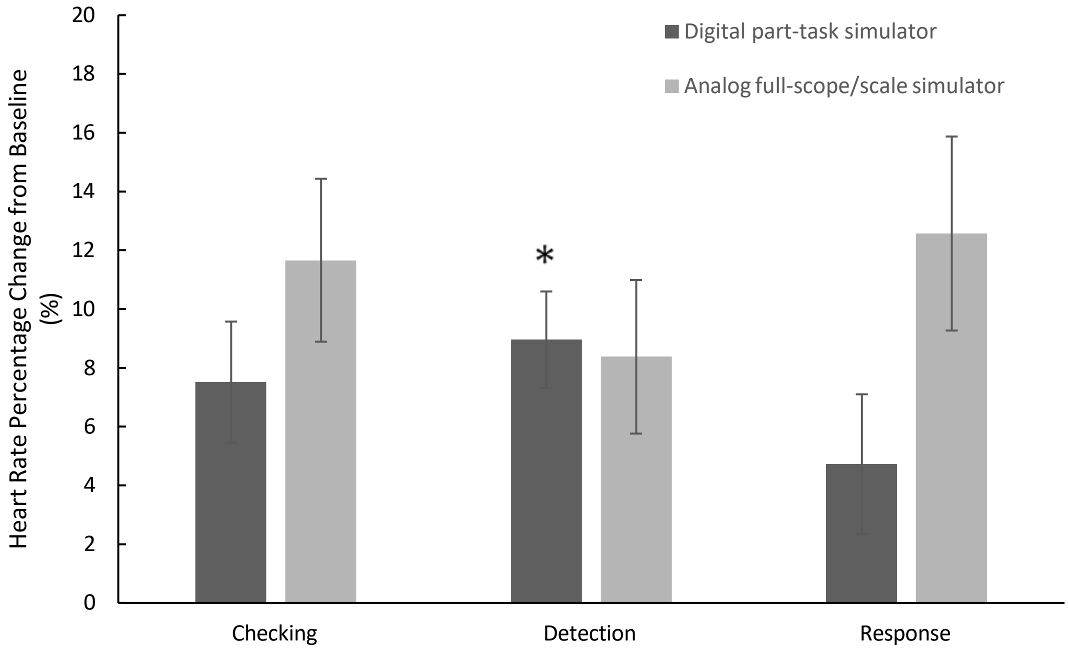

Figure 10 Heart rate percentage change from baseline by task type (error bars denote standard errors)

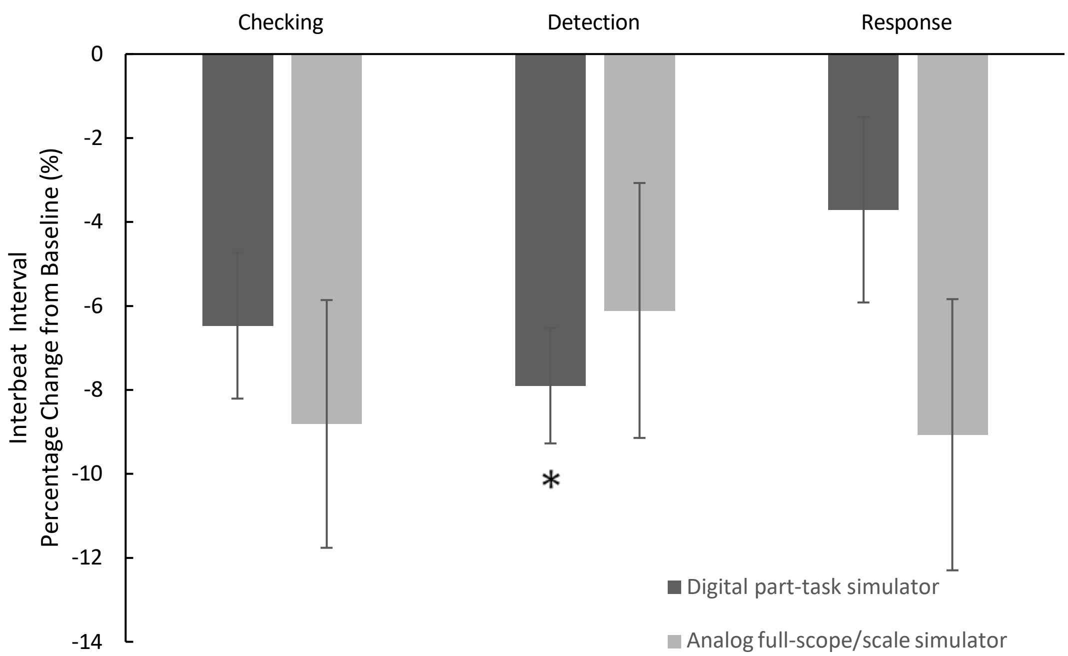

Figure 11 Interbeat interval percentage change from baseline by task type (error bars denote standard errors)

3.3 Performance Measures for Communication

Communication reporting variables included percent communications completed correctly, number of I&C location help requests, number of clarifications required, and number of requests for repeating an instruction. Three (Task Type: checking, detection, and response implementation) × 2 (Simulator: digital, part-task and analog, analog, full-scope/scale) × 2 (Operator role: RO and BOP) mixed ANOVAs with repeated measures on task type were conducted for each of the four measures to test the effect of task type and simulator type on communication between operator roles.

A significant main effect of task type was found for percentage of communications completed correctly, $F(1.25,50.16) = 5.14$, $p < .05$, $\eta_p^2 = .11$. Participants showed highest percent communications completed correctly in the response implementation task type ($M = 67.05$ $SD = 26.81$) than in the checking task type ($M = 56.81$ $SD = 29.98$) or the detection task type ($M = 59.89$ $SD = 34.10$). The interaction between task type and simulator type was also significant, $F(1.25,50.16) = 4.90$, $p < .05$, $\eta_p^2 = .11$. In the digital part-task simulator condition, the detection task type had the lowest percent communications completed correctly ratings, whereas in the analog full-scope/scale condition, the lowest percent communications completed correctly ratings appeared in the checking task type (Figure 19).

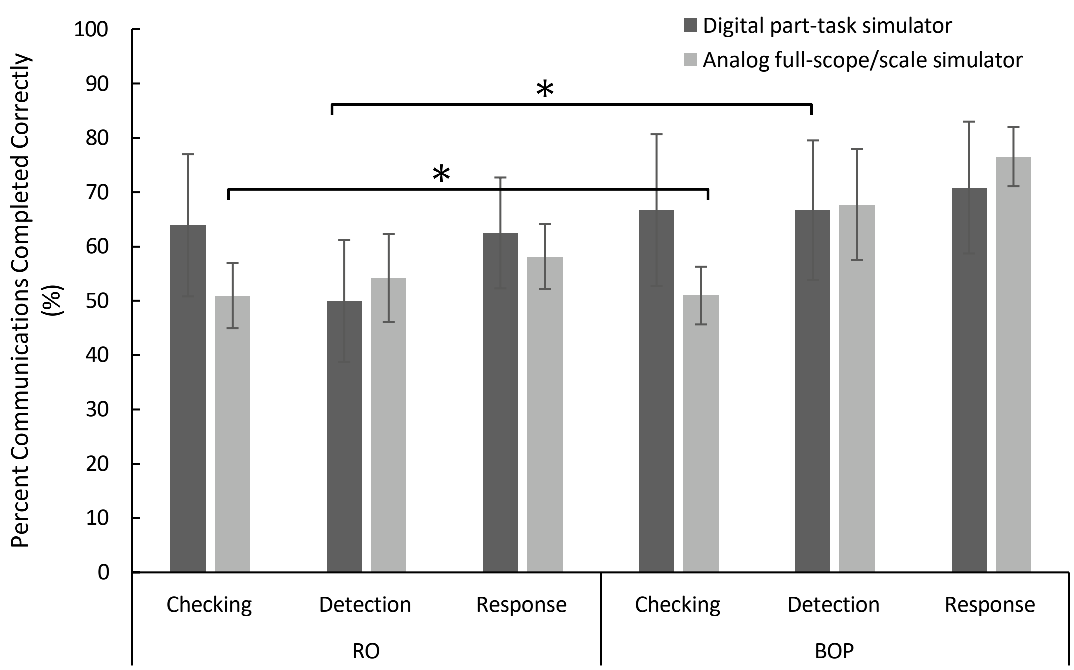

Figure 12 Percentage of communications completed correctly by task type and operator role (error bars denote standard errors)

For the number of clarifications required, the main effect of task type was significant, $F(1.29,51.61) = 4.52$, $p < .05$, $\eta_p^2 = .10$. Participants required more clarifications in the detection task type ($M = 7.68$ $SD = 16.30$) than in the response implementation task type ($M = 2.35$ $SD = 7.32$) (Figure 20). No other main effect or interaction was significant.

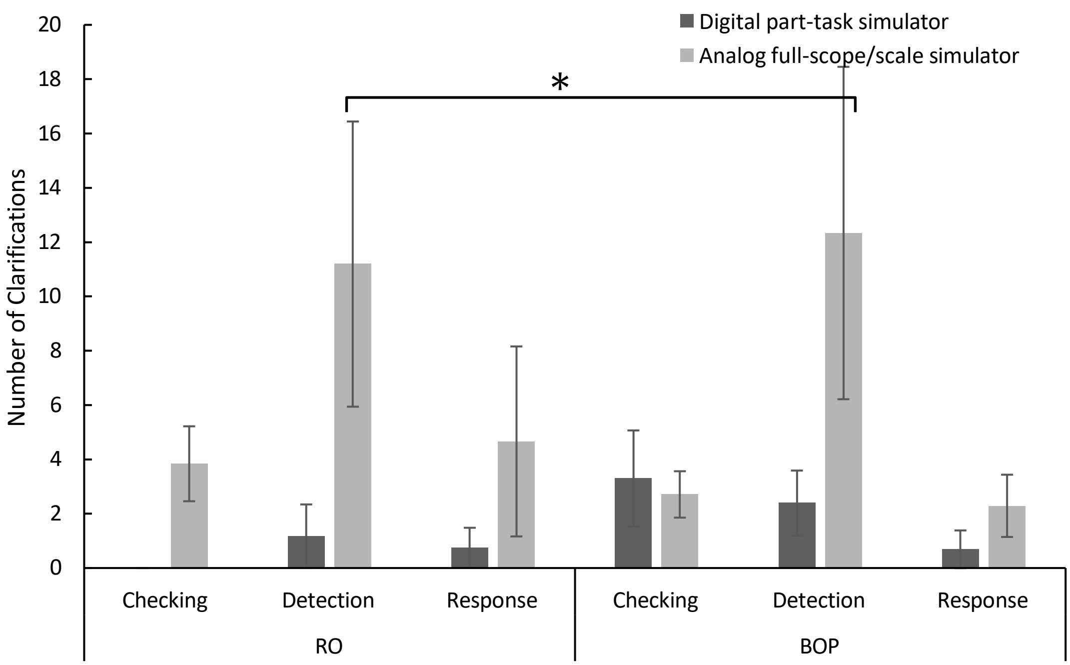

Figure 13 Number of clarifications by task type and operator role (error bars denote standard errors)

For the number of requests for repeating an instruction, the main effect of task type was significant, $F(1.71,68.24) = 10.61$, $p < .01$, $\eta_p^2 = .21$. Participants asked for more repeat requests in the detection task type ($M = 5.80$ $SD = 8.26$) than in the checking task type ($M = 2.55$ $SD = 3.72$) and the response implementation task type ($M = 2.77$ $SD = 5.03$). The main effect of simulator was also significant, $F(1,40) = 8.99$, $p < .01$, $\eta_p^2 = .18$. Participants in the digital part-task simulator condition ($M = 6.11$ $SD = 6.93$) asked for more repeat requests than the participants in the analog full-scope/scale simulator condition ($M = 2.04$ $SD = 5.76$). In addition, the interaction of task type and simulator type was also significant for the number of requests for repeating an instruction, $F(1.71,68.24) = 6.10$, $p < .01$, $\eta_p^2 = .13$. In the digital part-task simulator condition, participants asked for more repeat requests in the detection task type, but such trend in the analog full-scope/scale simulator condition was less noteworthy (Figure 21).

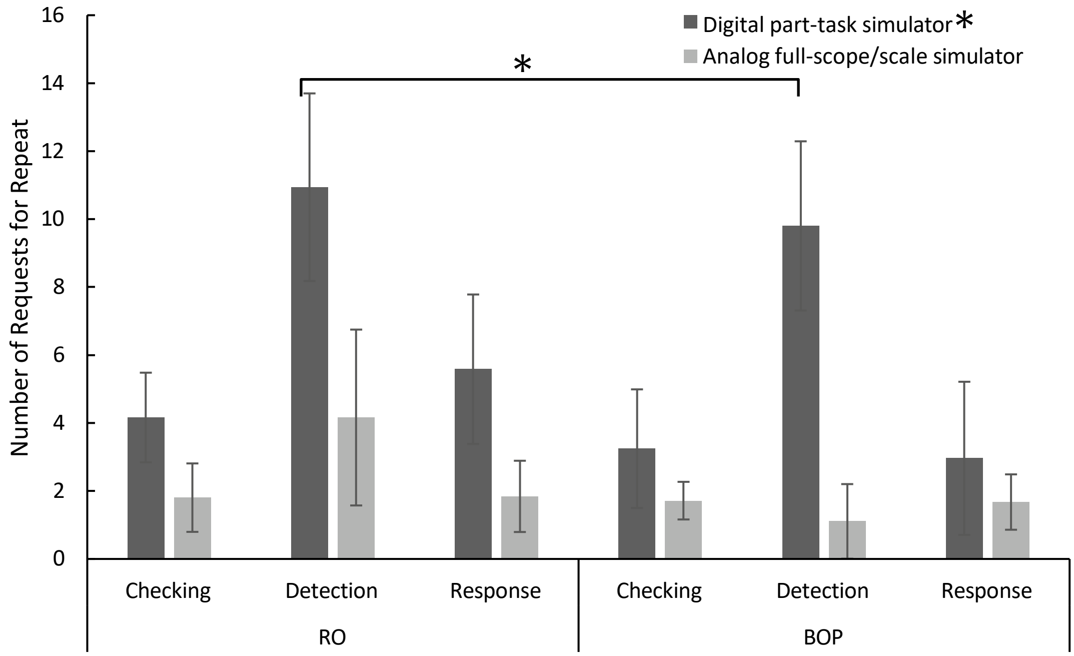

Figure 14 Number of requests for repeat by task type and operator role (error bars denote standard errors)

4 DISCUSSION

4.1 Subjective Measures

The subjective measures suggested that the analog, full-scope/scale simulator may elicit higher workload than the digital, part-task simulator, but only for select subscales. For example, the MRQ indicated that more mental processes in terms of spatial quantitative, visual temporal, and vocal are utilized in the analog, full-scope/scale simulator condition; the ISA also showed that participants perceived higher workload in the in the analog, full-scope/scale simulator condition. The experimental procedure used in the analog, full-scope/scale simulator condition had more steps and panels due to the nature of the different plants upon which the two simulators were based. Those additional steps and added complexity of the interface may impose extra workload, such as demand for spatial quantitative process and visual temporal process, upon operators. In addition, the communication analysis suggested that participants in the analog, full-scope/scale simulator condition had a trend of requesting more clarifications. Interestingly, although other measures suggested that the analog, full-scope/scale simulator may elicit higher workload, the NASA-TLX revealed significantly lower frustration in the analog, full-scope/scale simulator condition. This may indicate that operator with the experience in working with analog systems may get frustrated with a touchscreen interface. Adequate training will be necessary for operators working with a different interface.

The NASA-TLX results also revealed a similar pattern of interactions between simulator type and operator role, which supported the conclusion of detection task type as the most demanding one. Participants in the RO role in the analog, full-scope/scale simulator condition had comparable detection tasks and more overall steps, therefore, they reported higher ratings in several NASA-TLX subscales. On the contrary, participants in the BOP role in the analog, full-scope/scale simulator condition had few detection tasks, resulting in their low ratings. Participants in the BOP role, especially in the analog, full-scope/scale simulator condition, also required more movement to complete the assigned tasks and work with more panels. The significantly higher ratings in manual process and spatial attentive process measured by the MRQ confirmed participants in the BOP role perceived higher workload in those aspects.

4.2 Physiological Measures

The significant task type effect revealed by EEG and fNIRS data supported the conclusion of the detection as the most demanding task type from HPTF 1. EEG theta, beta, and alpha bands all showed significant differences in change during the detection task type session. Generally, compared with the checking and the response implementation task types, the detection task type elicited a smaller increase in EEG theta, beta, and alpha bands. Theta and beta increases are associated with workload (Kurimori & Kakizaki, 1995). Theta has been shown to increase with concentration and working memory demands. Beta is associated with arousal, attention, and workload. The smaller beta increase for the detection task type may indicate a loss of sustained attention and vigilance, even with no decrement in performance. Regional blood oxygen saturation (rSO_2) measured by fNIRS was highest for the detection task type which suggests a greater level of cognitive activities.

The EEG data also revealed a simulator effect. Generally, all four EEG bands showed a similar trend across task types and hemispheres/lobes, but the changes in the analog, full-scope/scale simulator condition were in a greater magnitude than in the digital, part-task simulator condition. Due to the layout of the analog, full-scope/scale simulator, locating specific I&C across panels to identify task components in the analog, full-scope/scale simulator condition may be more demanding than in the digital, part-task simulator condition. Previous studies suggested that the navigation and identification task components were the major source of working memory and concentration demands, however, note that the overall level of workload observed across all groups was fairly modest. This finding is consistent with our hypothesis. In addition, ECG metrics, HR and IBI, revealed significant difference for the detection task type in the digital, part-task simulator condition, but not in the analog, full-scope/scale simulator condition. The general cardiovascular pattern, such as increased HR and decreased IBI, in the analog, full-scope/scale simulator condition still supported increases in workload during the experimental manipulation.

4.3 Performance Measures for Communication

Participants in both conditions can communicate with the SRO effectively and execute the tasks assigned by the SRO with relatively few errors, even when skipping some parts, such as acknowledgement or confirmation, in the standard three-way communication. Participants showed the highest percentage of communications completed correctly in the response implementation task type. Participants were more likely to skip the acknowledgement or confirmation part when manipulation of an I&C was not involved in a step, such as the checking or detection task type. More requests for clarifications and repeating in the detection task type indicates that it is the most demanding task and may suggest that the detection task type could lead to confusion.

4.4 Future Directions

In order to get a more holistic perspective on the data previously collected in the HPTF experimental series, future work on the methodology should perform additional analyses on the data previously collected in the HPTF experimental series. The analyses resulting from this ‘deeper dive’ into the data will be published in subsequent volumes to this Research Information Letter series and will further examine:

- sensitivity of workload measures across the studies and the tasks (RIL 2022-11, volume 3)
- order effects of the tasks (RIL 2022-11, volume 4)
- potential ergonomic issues related to shoulder height and sensitivity of the touchscreens versus desktop interface control techniques (RIL 2022-11, volume 5)

4.5 Conclusion

This study investigated the impact of simulator type and task type on operators’ workload responses using the same methodology designed in the baseline HPTF experiments. Participants performed the same scenario following a similar, (modified) procedure using either an analog, full-scope/scale simulator, or a digital, part-task simulator. Although the simulator type had a significant effect in several metrics among subjective, physiological, and performance measures, such as NASA-TLX, ISA, EEG, and communication, generally similar trends were observed from the results. Due to the nature of the two simulators, the operating procedures were not identical, and the two operators in the analog, full-scope/scale simulator condition had unequal numbers of steps in terms of the three task types. Such differences in experimental scenarios and interface complexity of the two simulators (e.g., number of panels) may account for the differences in workload response between the two simulator conditions. Additionally, because of limitations of the analog simulator it was difficult to measure operator performance. The research team used communication performance because that data could be observed from video recordings and did not require the simulator to be capable of collecting behavioral data. Future studies with analog and digital simulator comparisons should address this limitation so that performance comparisons can be made between the two simulator types.

Overall, it is feasible to use digital simulators to conduct research and demonstrate trends in operator workload and performance between digitized and analog simulated environments. Working with digital simulators may enable the identification of safety concerns under circumstances where there is no easy access to the analog environment. Use of digital simulators may also benefit supplemental operator training. The results of the study reported in volume 2 support that research using digital simulators can be conditionally generalized to analog systems. The study also validated the methodology of using multivariate assessments to profile operator workload in the nuclear domain. NPP operations often involve workload variation and different types of tasks that may require different mental resources; therefore, it is difficult to use a single measure to profile workload comprehensively. Multivariate assessment of workload can provide a more detailed picture of operator’s neural and cognitive response to the task manipulations.

5 REFERENCES

Ayaz, H., Shewokis, P. A., Curtin, A., Izzetoglu, M., Izzetoglu, K., & Onaral, B. (2011). Using MazeSuite and functional near infrared spectroscopy to study learning in spatial navigation. *Journal of visualized experiments*, *56*, 3443.

Boles, D. B. (1991). Factor analysis and the cerebral hemispheres: Pilot study and parietal functions. *Neuropsychologia*, *29*(1), 59-91.

Boles, D. B. (1992). Factor analysis and the cerebral hemispheres: Temporal, occipital and frontal functions. *Neuropsychologia*, *30*(11), 963-988.

Boles, D. B. (1996). Factor analysis and the cerebral hemispheres: “Unlocalized” functions. *Neuropsychologia*, *34*(7), 723-736.

Boles, D. B. (2002). Lateralized spatial processes and their lexical implications. *Neuropsychologia*, *40*(12), 2125-2135.

Boles, D. B., & Adair, L. P. (2001). The multiple resources questionnaire (MRQ). *Proceedings of the human factors and ergonomics society annual meeting*, *45*(25), 1790-1794.

Chance, B., Zhuang, Z., UnAh, C., Alter, C., & Lipton, L. (1993). Cognition-activated low-frequency modulation of light absorption in human brain. *Proceedings of the National Academy of Sciences*, *90*(8), 3770-3774.

Harris, J. (2017). *Investigating the impact of levels of experience on workload during nuclear power plant operations* (Doctoral dissertation). University of Central Florida, Orlando, FL.

Harris, J., Reinerman-Jones, L. E., & Teo, G. (2017). The impact of simulation display on nuclear power plant task error frequencies. In S. Cetiner, P. Fechtelkotter, & M. Legatt (Eds.), *Advances in human factors in energy: Oil, gas, nuclear and electric power industries* (pp. 133-144). Cham, Switzerland: Springer.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock, & N. Meshkati (Eds.), *Advances in Psychology* (Vol. 52, pp. 139–183). Amsterdam, the Netherlands: North-Holland.

Henelius, A., Hirvonen, K., Holm, A., Korpela, J., & Muller, K. (2009). Mental workload classification using heart rate metrics. In *Engineering in Medicine and Biology Society, 2009: Proceedings of the Annual International Conference of the IEEE* (pp. 1836–1839). New York, NY: IEEE.

Hughes, N., D’Agostino, A., & Reinerman, 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.

Hugo, J.V. & Slay III, L. (2017). Human factors and modelling methods in the development of control room modernization concepts. *Proceedings for NPIC&HMIT 2017,* San Francisco, CA.

International Organization for Standardization (2013). *Ergonomic design of control centres – Part 4: Layout and dimensions of workstations* (ISO 11064-4: 2013).

Joe, J. C., & Boring, R. L. (2017). Using the Human Systems Simulation Laboratory at Idaho National Laboratory for Safety Focused Research. In S. Cetiner, P. Fechtelkotter, & M. Legatt (Eds.), *Advances in human factors in energy: Oil, gas, nuclear and electric power industries* (pp. 193-201) Cham: Springer.

Joe, J. C., Boring, R. L., & Persensky, J. J. (2012). Commercial utility perspectives on nuclear power plant control room modernization. *8th international topical meeting on nuclear power plant instrumentation, control, and human-machine interface technologies (NPIC&HMIT)*, 2039-2046.

Kurimori, S., & Kakizaki, T. (1995). Evaluation of work stress using psychological and physiological measures of mental activity in a paced calculating task. *Industrial health*, *33*(1), 7-22.

O'Hara, J. M., Brown, W. S., Lewis, P. M., & Persensky, J. J. (2002). *Human-system interface design review guidelines* (NUREG-0700, Revision 2). U.S. Nuclear Regulatory Commission. https://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr0700/

O'Hara, J. M., Higgins, J. C., Fleger, S. A., & Pieringer, P. A. (2012). *Human factors engineering program review model* (NUREG-0711, Revision 3). U.S. Nuclear Regulatory Commission. https://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr0711/

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.* United States Nuclear Regulatory Commission Rockville, MD.

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

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

Tattersall, A. J., & Foord, P. S. (1996). An experimental evaluation of instantaneous self-assessment as a measure of workload. *Ergonomics*, *39*(5), 740-748.

Taylor, G., Reinerman-Jones, L. E., Cosenzo, K., & Nicholson, D. (2010). Comparison of multiple physiological sensors to classify operator state in adaptive automation systems. *Proceedings of the human factors and ergonomics society annual meeting*, *54*(3), 195-199.

U. S. Nuclear Regulatory Commission. (1995). *NRC's object-oriented simulator instructor station* (NUREG-1527). Retrieved from https://inis.iaea.org/collection/NCLCollectionStore/_Public/26/077/26077364.pdf

U. S. Nuclear Regulatory Commission. (1998). *NRC Regulations Title 10, Code of Federal Regulations*. Retrieved from http://www.nrc.gov/reading-rm/doc-collections/cfr/part050/full-text.html

U.S. Nuclear Regulatory Commission. (2006). Staff Requirements – Meeting with Advisory Committee on Reactor Safeguards (SRM-M061020). U.S. Nuclear Regulatory Commission. https://adamsxt.nrc.gov/navigator/AdamsXT/content/downloadContent.faces?objectStoreName=MainLibrary&ForceBrowserDownloadMgrPrompt=false&vsId=%7b440AF45F-D515-452A-BD02-245777EF9548%7d

U.S. Nuclear Regulatory Commission. (2008). Analysis of options and recommendations for new reactor simulator training of NRC inspectors (SECY-08-0195). U.S. Nuclear Regulatory Commission. https://adamsxt.nrc.gov/navigator/AdamsXT/packagecontent/packageContent.faces?id={BE294C07-D4F8-4030-8884-E67C9939CD85}&objectStoreName=MainLibrary&wId=1669916626294

U.S. Nuclear Regulatory Commission. (2009). Staff Requirements – Briefing on risk-informed, performance-based regulations (SRM-M090204B). U.S. Nuclear Regulatory Commission. https://adamsxt.nrc.gov/navigator/AdamsXT/content/downloadContent.faces?objectStoreName=MainLibrary&ForceBrowserDownloadMgrPrompt=false&vsId=%7bA48838E6-D0A2-4969-B4CE-F168068567B6%7d

U.S. Nuclear Regulatory Commission. (2016). *Standard Review Plan for the Review of Safety Analysis Reports for Nuclear Power Plants: LWR Edition - Human Factors Engineering (NUREG-0800, Chapter 18)* (NUREG-0800, Chapter 18). U.S. Nuclear Regulatory Commission. https://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr0800/ch18/index.html

Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology*, *12*(1), 3–18. https://doi.org/10.1207/S15327108IJAP1201_2