Investigating the Impact of Levels of Experience on Workload During Nuclear Power Plant Operations

Jonathan Harris
University of Central Florida

INVESTIGATING THE IMPACT OF LEVELS OF EXPERTISE ON WORKLOAD DURING NUCLEAR POWER PLANT OPERATIONS

by

JONATHAN T. HARRIS
M.S. University of Central Florida, 2007
B.S. University of Central Florida, 2004

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Industrial Engineering
in the Department of Industrial Engineering and Management Systems
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Spring Term
2017

Major Professor: Lauren Reinerman-Jones

ABSTRACT

The human-machine interface (HMI) of a Nuclear Power Plant (NPP) Main Control Room (MCR) is complex. Understanding HMI factors that influence Reactor Operator (RO) performance and workload when controlling an NPP is important. The Nuclear Regulatory Commission (NRC) began a program of research known as the Human Performance Test Facility (HPTF) with the goal of collecting human performance data to better understand cognitive and physical elements that support safe control room operation. The HPTF team developed an experimental methodology to evaluate workload using perceived ratings, performance measures, and physiological correlates. This methodology focuses on tasks commonly performed during operations in an NPP. These tasks include monitoring plant parameters, following defined procedures, and manipulating controls to change the state of the NPP. O'Hara and colleagues developed a framework for task classification. Reinerman-Jones and colleagues modified this framework such that monitoring and detection are separate task types. The task types (i.e., checking, detection, and response implementation) selected for experimentation are composed of steps within defined operating procedures that are rule-based.

Testing workload using sufficient numbers of ROs is impractical due to limited availability. The HPTF has developed the "equal but different" principle. This principle attempts to simplify complex tasks, such that novices can perform them and experience equivalent workload trends as an expert would when performing the original task. The validity of using the "equal but different" principle with novices in place of experts is uncertain. This research addresses this uncertainty by comparing novices and experts using the "equal but different"

principle. Novices performed four tasks within each of the three task types using a simplified Instrument and Control (I&C) panel and a reduced 3-way communication instruction set. Experts performed the same four tasks within each task type with a fully configured I&C panel and a complete 3-way instruction set.

Overall, the experts across the three task types tended to rate level of perceived workload lower than novices. However, experts also rated themselves as performing worse for the three task types than novices. Experts performed better than novices when it came to identifying correct I&C; however, their 3-way communication performance was worse. Physiological measures from EEG between the two groups were not statistically different. ECG findings did show a slight difference.

The methodology and associated findings has applicability for MCR designs and regulation recommendations. Novice populations are easier to access than experts and the present research shows that when properly designed, novices can serve in complex operator positions.

This work is dedicated to the love of my life, Sarah and my two girls Hadassah and Hannah.

ACKNOWLEDGMENTS

This work was supported in part by the United States Nuclear Regulatory Commission. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of U.S. Nuclear Regulatory Commission or the U.S. Government.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

The commercial nuclear power industry accounts for around 20 percent of energy produced in the United States (IAEA, 2016). Nuclear Power Plants (NPP) are operated with relatively low accident rates of 0.13 events per 200,000 worker-hours (Heiser, 2009). However, the outcome from any NPP event can be devastating. The NUREG 75/014, Reactor Safety Study has estimated NPP core melt accidents at the rate of one in 20,000 per reactor per year (U. S. N. R. Commission, 1975). The Main Control Room (MCR) is where the operators interface with the displays and controls of the NPP. The present study views NPP MCRs as both complex and safety-critical systems.

There are numerous definitions in almost every area of research for what comprises a complex system. No generally agreed upon universal definition for complex systems exists; for the purposes of this study, the criteria for a complex system will be a system made up of a large number of sub-systems (i.e., components) with interworking relationships among them, such that the behavior of each sub-system depends on the behavior of other sub-systems (Simon, 1962). Knight (2002) defines safety-critical systems as those necessary for ensuring public safety and health, and the wellness of the environment and plant workers.

Numerous studies have looked into the root causes of accidents in industrial organizations. Accidents in industries such as aviation, rail, and manufacturing facilities have found that human error plays a significant role in approximately 30 percent of accidents. Similar statistics can be found in NPP operations (Park, Jung, & Yang, 2012; Seong et al., 2013). This statistic highlights the importance of understanding the role humans play in maintaining safe and efficient operations in an NPP MCR.

The historical NPP meltdown at Three Mile Island Unit 2 (TMI-2) was caused in large part by human error. Out of the eight significant factors that led to the meltdown, five were related to human error (John, 2008). Two other factors were related to a design flaw and only one significant factor was due to equipment failure, which alone would not have been sufficient to cause the partial meltdown (John, 2008). The TMI-2 cleanup took 14 years and costs were estimated at $1 billion dollars ("14-Year Cleanup at Three Mile Island Concludes," 1993). TMI-2 was only a partial core meltdown where no lives were lost and the incident only had negligible impacts on human health and the environment (NRCgov, 2013). Nevertheless, the TMI-2 incident brought about sweeping changes in the way the Nuclear Regulatory Commission (NRC) regulates and licenses NPPs. These changes primarily impacted “human factors” in plant layout designs, human performance standards, and fitness for duty programs ("Three Mile Island Accident," 2013). After TMI-2, regulators adopted symptom-based Emergency Operating Procedures (EOPs) as a method to reduce workload for human operators and improve overall safety of the NPP under off-normal conditions.

The NRC is responsible for regulating guards against human-caused errors, such as those directly tied to the events of TMI-2 many years ago (NRCgov, 2013). In the United States, NPPs with digitally controlled MCRs are coming online within the next few years (Harris, Reinerman-Jones, & Teo, 2017). These new digital plants confer benefits such as improved automation over conventional analog controlled plants; however, these digital designs pose new human factors risks. In addition to the new plants, older plants are being modernized, creating a hybrid analog/digital MCR, which presents yet another set of human factors risks to address (Joe, Boring, & Persensky, 2012).

Nuclear Power Plant Main Control Room Operations

A Reactor Operator (RO) is a person who directly interacts with the HSI of the NPP MCR. The Skill-Rule-Knowledge (SRK) human behavior classification framework (Rasmussen, 1983) is a model that researchers have used to aid in understanding interactions between the RO and the HSI (Lin, Yenn, & Yang, 2010). Crew coordination through effective teamwork and efficient communication is vital to MCR design (Fink, Hill, & O'Hara, 2004). An RO is a member of the crew that operates the NPP with a common goal of safe and efficient power generation. To achieve this, it is necessary to understand (i) the tasks involved, especially (ii) how the crew communicates, coordinates, and executes the tasks within the complex system, which includes (iii) the system design and layout of the MCR.

Main Control Rooms

All details described herein are with respect to NPP MCRs in the United States. The MCR is where all the Instruments and Controls (I&C) are housed that control the reactor and associated safety systems. The MCR boundaries are specified as the “vital area” and are defined in Title 10 of the Code of Federal Regulation (CFR) section 73.2. MCR controls are defined in 10CFR 50.54 as any apparatus that directly affects the reactivity or power output levels of the nuclear reactor (N. R. Commission, 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 a MCR crew is dependent on the number of reactor units and control rooms at the NPP site and are defined in 10 CFR 55.54(2)(i).

Typically, a NPP crew is made up of at least one licensed Senior Reactor Operator (SRO) in the control room, as well as one licensed RO present at the controls at all times and one or more relief operators licensed and able to take on the role of operator at the controls. The SRO assigned to "control room duties" are required to be within eyesight or audible range of the operators at the controls (U. S. N. R. Commission, 2008). SRO duties are supervisory in nature. SROs are stationed in the MCR where they have direct and prompt access to information on the current state of the plant. The SRO should maintain situational awareness of the plant's state, provide expertise and knowledge in the event of an off-normal condition occurring, and execute emergency procedures in the event of multiple alarms or a reactor trip.

The RO is primarily tasked with insuring that the reactor unit is operating safely. The RO at the controls is required to stay within the surveillance area of the MCR with an unobstructed view of the operational control panels and annunciators. Relief operators are also licensed ROs typically tasked with aiding the operator at the controls (U. S. N. R. Commission, 2008).

Layout

An efficient, reliable, and consistent HSI is the ergonomic goal for control room design (Raeisi, Osqueizadeh, Maghsoudipour, & Jafarpisheh, 2016). NPP MCR designs have been developed and modified over many years; as a result, most NPP MCR I&C layout and workstation configuration is unique. However, each operational MCR is required to have 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.

MCR workstations are where ROs perform their tasking. They contain HSI elements that control normal operations of the plant and associated safety systems at the plant. Workstation types traditionally found in NPP MCRs are standup-consoles, sit-down consoles, sit-stand workstations, and vertical panels (O'Hara, Brown, Lewis, & Persensky, 2002). Ergonomic factors such as control location, visual layout, and overall comfort all affect RO performance and workload.

A MCR contains workstations and other equipment (e.g., spare parts, tools, emergency equipment such as protective clothing, etc.), as well as documentation (e.g., safety procedures and manuals). Ergonomic configuration of the control room is determined by the arrangement of workstations, proper storage and location of equipment, and the organization of document storage for ease of access (O'Hara et al., 2002).

The layout characteristics of the MCR workspaces directly affect the functionality of the HSI. The 100 operational NPPs in the United States have MCRs consisting of analog I&C workstations or hybrid (analog/digital) workstations (Joe & Boring, 2017). Fully digital MCRs are coming online in the next few years (Harris et al., 2017). Differences in workstation configurations impact workload associated with the HSI in different ways. These HSI differences not only impact individual crewmember tasking, but also teamwork cohesion (Fink et al., 2004).

Proper understanding and assessment of the impact the HSI has on crewmember tasking is important for a proper MCR design.

Communications

The SROs and ROs work as a team with the common goal of operating the NPP in a way that maintains safe and efficient power generation. As teamwork is essential for effective and safe nuclear operations, the SROs and ROs in the same team typically share the same work, training, and rest schedules (Joe & Boring, 2017) to promote coordination and communication within the team. Poor communication has been regarded as one of the main causes of team coordination issues and affect team performance and task quality (Kim, Park, Han, & Kim, 2010). Billings and Cheaney analyzed 28,000 incident reports in NASA's Aviation Safety Reporting System and found that voice communication issues were present in over 70% the incidents.

The goal for communication during the tasking of ROs is to convey information accurately so that there is common understanding among the team members. As this communication often involves technical information related to the safety of the plant, it is paramount that all crewmembers understand the information correctly (Min, Chung, & Yoon, 2004). Most of the communication among team members is in the form of oral communication. However, although oral communication is less time-consuming and allows quicker transfer of information, it is more susceptible to misunderstanding and various inaccuracies compared to written communication.

In order to facilitate effective oral communication, operators in U.S. NPP MCRs utilize a repeat back method known as three-way communications (DOE, 2009). Three-way communication helps ensure the reliable and accurate transfer of information between two people. A typical three-way message starts with the first communication being a crewmember addressing another by name and issuing a short instruction. The second part of the three-way message has the addressee of the message echoing back the instruction that was understood by the addressee in a paraphrased manor. The paraphrased instruction must contain the technical details of the instruction. Should the addressee need clarification, more detail, or did not understand the instruction, rather than echoing back the message, the addressee would request the needed information via a repeat request. Once the addressee has correctly echoed back the instruction, the initiating crewmember closes the loop by affirming that the instruction was understood correctly. Studies have found a positive correlation between the use of three-way communication and performance (Kim et al., 2010).

RO Task Types

ROs control the NPP through the HSI. O'Hara and colleagues have developed a framework for task classification for interacting with the HSI (O'Hara & Higgins, 2010). Their framework bifurcates RO tasks into primary and secondary tasking. Secondary tasking are tasks that deal with interface management such as navigating through or accessing information at a workstation. Understanding secondary tasking is important in the context of primary task performance because success of the primary task is dependent on information retrieval of the secondary task. Interface tasks such as navigating, locating, and arranging information at a workstation are categorized as secondary. This is because these tasks are not directly associated with monitoring or controlling the plant (O'Hara & Higgins, 2010).

Primary tasks are associated with controlling and running the NPP and include monitoring plant parameters, following defined procedures, as well as manipulating controls to change the state of the NPP. O'Hara and colleagues defined four distinct elements of primary tasks: Monitoring and detection, assessing situations, planning responses, and response implementation. Reinerman-Jones and colleagues follow similar classifications; however, they regarded monitoring and detection as two separate task types: the checking and detection tasks (Reinerman-Jones, Guznov, Mercado, & D'Agostino, 2013). The checking task involves a discrete one-time assessment of an I&C to verify its current state or level (e.g., verify that a certain valve is open). It requires observing readings on the displays, viewing I&Cs as well as processing verbal reports from other team members. The checking task type is a successive-attention task where multiple checks are performed back to back. The successive-attention component of the checking task maintains consistent task demands throughout the task and requires operators to retain critical information in their working memory and distinguish an indicator from a non-indicator (Reinerman et al., 2006).

The detection task is a continuous task where an RO is required to monitor state changes as reflected in the I&Cs and to report back when a certain state has been reached (Reinerman-Jones et al., 2013). This task type stems from signal detection theory, which requires participants to remain vigilant and to discriminate noise from signals with noise (Tanner & Swets, 1954). In the detection task type, operators would be watching a gauge's level change. Two factors influence the operator's ability to detect gauge level changes, (1) signal to noise sensitivity and (2) detection bias. Sensitivity has to do with the ability of the operator to discriminate the signal from the noise (Wickens, Hollands, Banbury, & Parasuraman, 2015), and is influenced by the operator’s ability to remain vigilant at detecting changes over a period of time. Bias impacts an operator’s preference for erring on over/under reporting changes. Over reporting bias indicates that the operator is cautious. Thus he/she would rather report a change when one doesn’t exists rather than miss any changes.

Planning response tasks deal with optimally controlling the NPP in a safe state. In the NPP domain, planning responses is typically done offline using symptom-based Operating Procedures (OPs). These include Emergency Operating Procedures (EOPs), maintenance procedure, and daily operations. OPs are step by step rule based procedures defining the appropriate actions to perform on the NPP. The goal of OPs are to maintain the NPP in a safe state, or bring the plant back to a safe state optimally in the event of an off-normal event.

The response implementation task type requires the RO to take action on the I&Cs to modify the state of the NPP (e.g., shutting or opening a valve). Response implementation can occur through direct wired analog controls and through digital soft controls found in modern MCRs. With both analog and digital controls, operators most use fine motor skills in order to rotate switches or valves to impact the state of the NPP.

Understanding RO Workload during NPP MCR Operations

Safe Control Room Practices

The office of the Nuclear Regulatory Commission began a program of research known as the Human Performance Test Facility (HPTF) with the goal of collecting empirical human performance data to aid in understanding cognitive and physical elements that support safe control room operation (Hughes, D’Agostino, & Reinerman-Jones, 2017). There have been many studies in the past looking at NPP MCR operations through human reliability analysis (HRA).

However, NPP HRA is difficult to quantify because there are large individual differences in the type and frequency of human errors, as well as various issues arising from HSI factors (Lee, Kim, & Jang, 2011). Understanding the impact on operator workload and performance is a widely studied area of human factors, with over five hundred workload articles in print since the 60s (P. A. Hancock & Meshkati, 1988). However, workload research in the nuclear domain has been limited (Reinerman-Jones, Hughes, & D'Agostino, 2016). In order to develop meaningful regulation, the NRC needs to better understand levels and types of RO workload elicited while performing different tasks common to MCR operations within specific HSI (Reinerman-Jones et al., 2016).

Workload Measures

Most agree that workload is multidimensional and is a result of the demand imposed by the task on an operator's mental resources. However, in the literature there are many conflicting ideas, definitions, and ways to measure workload (Moray, 2013). For the present study, workload is defined as "the operator's perceived evaluation and accompanying physiological response to the experience imposed by the task demands rather than a direct reflection of the task demands themselves" (Abich, 2013). Since the 1960s, the measurement of workload has been a significant area of research (Estes, 2015). By the late 1970s, researchers started measuring workload through subjective opinions, spare mental capacity tasks, and primary performance task measures as well as physiological correlates (Williges & Wierwille, 1979). The present study recognizes the importance of using multiple measures as a way to assess the multidimensionality of workload.

Self-report measures are the most widely used tool to assess workload, likely because of their ease of usability and face validity (Estes, 2015). The most commonly administered self-report tool used is the NASA-TLX (Hart & Staveland, 1988). NASA-TLX is referenced in more than 6000 published works including over 550 reviews of the tool itself. The tool measures six relatively independent subscales: mental, physical, temporal demands, frustration, effort, and performance. NASA-TLX is the most often administered post-task which requires operators to recall events. On the other hand, the Instantaneous Self-Assessment (ISA) is an online measure developed by the United Kingdom Civil Aviation Authority as a simple immediate rating of work demand during primary task execution(Tallersall & Foord, 1996). The ISA is administered with a short auditory prompt that signals the operator to rate his/her current global workload on a 5-point Likert scale ranging from being under-utilized to experiencing excessive workload. The present study will utilize both the NASA-TLX and ISA as a means to assess the subjective facets of workload.

Task demands can be defined by the goals to be achieved. It can be evaluated by the time allowed to perform the task and the performance level to which the task is to be completed (Gawron, 2008). Performance measures provide a link between task demands and workload. Workload can remain constant with an increasing level of task demand if performance is allowed to decrease proportionally. However, performance alone is a poor indicator of workload because under certain conditions, dissociation between workload and primary task performance has been observed (P. Hancock, Williams, & Manning, 1995; Matthews, Reinerman-Jones, Wohleber, Lin, Mercado, & Abich IV, 2015; Matthews, Reinerman-Jones, Wohleber, Lin, Mercado, &

Abich, 2015; Mercado, 2014; Joseph E. Mercado, Lauren Reinerman-Jones, Daniel Barber, & Rebecca Leis, 2014; Yeh & Wickens, 1988).

Physiological measures of workload have advantages over subjective and task performance measures. They are objective and allow continuous monitoring of workload throughout the task. However, technical expertise is extensively required for analysis because there is no standardized scoring procedures (Kramer, 1991). This is due in part to variations in physiological response patterns. It has been observed that individuals produce different physiological responses to identical circumstances (Turner, 1994). In addition to individual differences issue, task types themselves produce different patterns (Miyake, 2001). Still with these limitations, ISO 10075-3:2004 “Ergonomic principles related to mental workload” has recognized the importance of incorporating physiological indices in their workload measurement method (10075-3, 2004).

Electroencephalography (EEG) has been used in numerous workload studies due to significant finding of EEG correlates with workload (Berka et al., 2007a). EEG is a direct measurement of activity in the central nervous system. It measures the electrical activity of the brain through electrodes placed along the scalp of the operator. Power spectral density analysis yields the theta, alpha, and beta frequencies, which have been shown to be sensitive to changes in workload (Berka et al., 2007b; Hankins & Wilson, 1998; Kurimori & Kakizaki, 1995; Wilson & Eggemeier, 1991). Theta, specifically in the frontal lobes and along the midline, during mental concentration tasks, is associated with a high-amplitude (Kubota et al., 2001). Similarly but in an inverse relationship alpha tends to decrease as workload demands increase (Gevins, Zeitlin, Doyle, Schaffer, & Callaway, 1979). Additionally, alpha has been shown to attenuate as visual scanning task complexity is increased (Gundel & Wilson, 1992). Wertheim's research shows the suppression in alpha during visual scanning tasks is caused by retinal involvement and oculomotor control (Wertheim, 1981). Beta activation has been shown to correlate with cognitive and emotional processing (Gundel & Wilson, 1992). Specifically, beta has been shown to increase with increases in arousal, attention and workload (Prinzel, Freeman, Scerbo, Mikulka, & Pope, 2000).

Electrocardiogram (ECG) is a measure of the electrical activity of a heart. ECG is one of the most frequently used physiological measure of workload (Mercado, 2014). Heat rate variability (HRV) and inter-beat interval (IBI) have been used to index workload. The general cardiovascular pattern for increases in workload are characterized by decreases in HRV and IBI (L. Mulder, D. de Waard, & K. A. Brookhuis, 2004).

HSI Modifications

According to the International Atomic Energy Agency (IAEA), currently sixty NPP are under construction around the world (Larson, 2016). There are five new Generation III reactor designs each with a different modernized MCR design. These new MCR designs have fundamentally changed the HSI for the RO. Many of the 100 U.S. commercially operated NPPs have gone through some partial modernization of the NPP MCR. However, none have completed a full control room modernization effort (Joe et al., 2012). These modernizing efforts directly impact the HSI and include changes to alarms, displays, and I&Cs in the MCR (Fink et al., 2004). Digital systems have been adopted in other critical process control domains, but advanced digital I&Cs in the nuclear domain is largely untested (Joe et al., 2012). The same gap in knowledge regarding the modernized control room's impact on workload also exists in hybrid analog-digital control rooms.

Abich found that workload measures change across task type (Abich, 2013). Automation, display layout, and soft control changes in modern digital MCRs have fundamentally altered the taskload of operating an NPP. In an HSI interface comparison between modern seated workstation and a simulated conventional standing touchscreen workstation, Reinerman and colleges (2016) found significant differences between both subjective and performance measures of workloads when performing identical tasks. The Reinerman study suggested that these different interfaces imposed different taskload, and their effects on workload experienced by the ROs is not fully understood. Any effort to improve understanding of this relationship between HSI and workload is especially important in emergencies where automation is often less effective and the cognitive demand of the operator is high. To that end, it is important to note that Reinerman et al. (2016) only reported using a novice population for two different interfaces. It is plausible that those results will generalize to an operator population because the question at hand dealt with a more traditional usability question pertaining to levels and types of workload and an expert population might show negative transfer when utilizing new interfaces.

Existing NPP MCR Operator Research

For the purposes of understanding workload in the NPP MCR domain, it is often not feasible to perform experiments using a large numbers of ROs (Leis, Reinerman-Jones, Mercado, Barber, & Sollins, 2014). That has lead research in the area of human factors to focus on recommendations by Subject Matter Experts (SME), industry questionnaires, and small sample studies with ROs. These research methods are good for uncovering facts, adapting theory and guiding future research; however, they lack ability to make statistical inference or support system validation (Ha, Seong, Lee, & Hong, 2007).

For instance, a review of the current literature on workload in the nuclear domain found flawed designs that limited the generalizability of study results to different NPP MCR tasks (Mercado, 2014). This led Mercado to propose an approach that experimentally tested the common task types during MCR operations, balanced task types with all I&C types used, and include used a large sample of participants to improve statistical inference (Mercado, 2014). While Mercado’s design was sound and more generalizable than the other NPP MCR workload research, he utilized novice participants. Hence, it is not known if study findings generalized to operational environments with highly trained ROs.

Plant designs, safety systems, and technology changes over the years have created an environment where the HSI for each NPP MCR is unique. Currently there are two main types of NPPs operating commercially in the Unites States: boiling water reactors (BWR) and pressurized water reactors (PWR). About a third of the 100 commercial reactors are BWR. A BWR heats water to steam, which powers the turbine. The remaining two-thirds are PWR. In a PWR, pressurized water is heated under high pressure to below the boiling point in the reactor. The pressurized water is circulated next to a lower pressure water system where it transfers its heat. The low pressure water turns to steam and powers the turbine.

While commercial and military reactors use the same reactor technologies, light box and status indicator norms differ. In commercial NPPs a red light indicates a valve or switch is open while in the military domain, red indicates the valve is closed.

Differences between plant types, designs, and indicators all impact workload related to the HSI. These differences highlight the importance for workload experimentation that is generalizable across task types and robust to potential HSI confounds.

Experience Level and Workload Assessment

Modernized MCR deigns have shifted the paradigm for control room designs. Legacy plants mainly rely on direct manipulation of analog controls and hardwired status indicators placed throughout the MCR. Modernized NPPs have shifted to soft controls, digital displays and seated workstations.

Novice ROs Workload during NPP MCR Operations

A number of factors often hinder the use of ROs and SROs for Workload assessment in the NPP domain. Training simulators are often booked to capacity to support operator training, leaving little to no time for experimentation. Also, since each simulator is required to maintain identical configuration to its paired operational MCR, modifying the simulator introduces negative training risks (Joe & Boring, 2017). In addition to logistical factors, the expense of obtaining a large sample of ROs for the duration needed to assess workload is impractical (Leis et al., 2014).

To address research gaps in understanding levels and types of workload for NPP MCR operations, a systematic research approach that balances real-word RO tasking with experimental repeatability and control is needed (Reinerman-Jones et al., 2015). Research studies need to be conducted with sample sizes large enough to provide sufficient statistical power to make meaningful inferences. One cost effective solution is to incorporate large numbers of novice participants in addition to small samples of ROs in experiment designs. A major drawback of novice participants is the lack of domain knowledge. However, the lack of domain knowledge has the benefit of unbiased assessment of the process and designs being evaluated. This is an important attribute especially when testing large paradigm shifts in system designs.

Levels of Proficiency

Figure 1: The Dreyfus Model of Skill Acquisition

The Dreyfus model of skill acquisition (Figure 1) provides a framework for categorizing a person's level of proficiency (Dreyfus & Dreyfus, 1980). The model splits the continuum from novice to expert in to five stages. Novice behaviors are classified as being able to adhere to taught rules or plans. At the experienced level, people have some perception of how actions relate to the goals. The experienced level requires both training and practice. The expert level is achieved through years of real world operational experience. At the expert level, people can intuitively grasp the situation based on a deep understanding of the system.

Rasmussen's human behavior classification of Skill–Rule–Knowledge (SRK) levels provides a method for task analysis based on levels of proficiency. Through the SRK lens, a complex task such as a team of operators controlling a NPP can be distilled to a list of sub-tasks. If an experimental design abstracts or modifies this sub-task list to exclude knowledge based tasks, then novice or experienced participants may be suitable proxies for expert ROs. Specifically, rule-based tasking is a prime candidate to investigate the validity for using novices as proxies for expert ROs. Rule-based tasks that can be administered as step-by-step instructions align with the Dreyfus model's capabilities of a novice. In off-normal NPP MCR operations, symptom-based EOPs are followed (Kim et al., 2010). Symptom-based EOPs consist of skill-based and rule-based tasks.

Holistic View of Workload

Just as a battery of measures is recommended for assessing workload for a particular task type, using different populations throughout the Dreyfus model provides holistic understanding of workload for a task type. Access to novice ROs is easier and cheaper than experienced and expert ROs. However, studies that utilize experienced ROs provide a more robust understanding of workload than do novices. Experts ROs provide the most lucrative insight in to workload because this is the desired level of proficiency for operational ROs. Nevertheless, because experts have domain knowledge, transfer effects can influence the effectiveness of workload measures. This is most prevalent when assessing designs that similar but change or challenge conventional paradigms (Novick, 1988). In addition to transfer effects, testing new nonoperational procedures on operational ROs has the potential of negative training.

Developing a Research Approach

The research methodology approach for using an NPP simulator outlined in (Reinerman-Jones et al., 2013)) is the paradigm used in this research effort. Using a mock MCR, real-world OPs can be simulated in a repeatable manner. Moreover, an OP best suited for non-operator population can be selected for use with novice or experienced populations. Selection of an appropriate OP should consider the following criteria. First, the OP should resemble the task flow that operators most commonly face. This will help eliminate atypical tasks and aid in achieving generalizable results. Second, the OP should include equitable tasking for the MCR team members to maintain ecological validity. Third, the OP should perform an equal ratio of task types. By maintaining this ratio, comparisons can be made across task types. Lastly, the OP should incorporate the major categories of I&Cs such that the experimental results are not specific to a single control type (Reinerman-Jones et al., 2013).

Equal but Different

The “equal but different” principle (Lackey, 2014) lays out a strategic plan for utilizing novice populations. The principle utilizes five strategies: (1) proper experimental design, (2) distill skills into core components, (3) scaffolding, (4) proficiency testing, and (5) interpreting results. Each of these strategies is explained below.

Proper experimental design starts with a clear research question. The researcher should focus the experimental design in real world conditions, but shape it through the lens of theory. The researcher should utilize SMEs with domain knowledge when designing the experiment to help structure real world tasks around theory-focused concepts. The core idea is to have a theoretical foundation and still maintain task realism.

Distilling tasks into components requires a cognitive task analysis. The researcher should use information theory (Shannon, 2001) as a theory driven method for equating cognitive demands between the target expert population and the novice population. Shannon equates information as a reduction in uncertainty (Cole, 1993). Thus, by distilling the task to core components and removing ancillary or irrelevant components, the researcher is providing novice populations with information. The aim for this strategy is to create equal cognitive demands between experts and novice populations for the core components under investigation.

The scaffolding strategy is founded on Lev Vygotsky's zone of proximal development theory (Vygotsky, 1987). Scaffolding in this context is developing skills and knowledge by instructor led learning. The instructor provides enough assistance to the novice so they can perform a basic task, but no more than is required. As the novice learns and develops the necessary skills for the basic task, the instructor increases task complexity by including more components. This process is repeated until all components of the core task are trained. Upon completion of the training, the scaffolding (instructor assistance) is removed. The instructor should not provide guidance during experimentation.

Throughout the scaffolding process, the instructor should provide systematic proficiency tests to the novices. Those tests guide the novice and instructor on the need for additional instructions and identifies the skills that have been learned. These proficiency tests should resemble the experimental tasks; however, they should also be distinct from experimentation scenarios to avoid priming effects.

If a researcher follows the four strategies above, then conclusions he/she draws from the novice population should be representative of the expert population. It is important not to overgeneralize results. Conclusions drawn from the original intent of the study are the most valuable.

Applying Equal but Different to NPP MCR Operations

The "equal but different" principle was used to abstract the MCR tasks that require expert operators to execute. The abstracted tasks are designed to be equivalent tasks that can be performed by a novice or experienced operator with limited training. The strategies outlined in “equal but different” were used to distill tasking in such a way that along the dimension of workload, novices experience equivalent workload as experts performing the original task.

Rarely do ROs perform knowledge-based tasks during operations in an NPP MCR. This is because tasks are driven by defined procedures. Most complex domains rely on procedure driven tasks to reduce human error (Mercado, 2014). During off-normal conditions, NPP MCR operations are dictated by rule-based EOPs. These procedures include tasks types such as checking, detection, and response implementation. While these EOPs are heavily rule-based procedures, domain knowledge and skill are still required to execute them. Instructions in the EOP often use acronyms and names familiar to people who work in the industry, but are completely foreign to an outsider. The “equal but different” strategies are used to reduce this gap through simplification of the instruction, and the removal, exchange or clarification of ambiguous terms. Working memory limits of 7 plus or minus 2 items (Miller, 1956) is incorporated in the simplified instruction. This helps to eliminate working memory confounds in the instructions. The “equal but different” principle further mitigates the knowledge gap through targeted training of the specific instruction types encountered in experimentation.

The SRO is responsible for executing the OP. The SRO delegates tasks to ROs through written or verbal instructions. The specific lexicon for verbal instructions are unique to each MCR. However, most operating crews follow the paradigm that instructions should include a system name and the I&C specific alpha numeric identifier for locating an I&C. The system name provides context for locating the I&C such as the panel on which it is located. However, the wording of the instruction can vary significantly. An Instruction such as “verify” can have drastically different meanings. At one plant, a request to verify a valve is closed can mean to visually inspect the valve and to report back the current state, but not to close it if it is open. However, at another plant it could mean to visually inspect the valve and close it if it was in an open state.

The “equal but different” principle used for communication is to adopt a standardized base lexicon using SME input. Given this, the objective is to simplify verbal instructions further by referring to I&Cs only by their alpha numeric nomenclature and avoid obfuscating the instructions with meaningless system names for novices.

RO workstations are visually complex with multiple primary and safety system panels, each with hundreds of I&Cs on them. Research has shown that experts are better at distinguishing relevant vs irrelevant information in visually complex environments (Jarodzka, Scheiter, Gerjets, & Van Gog, 2010). Because experts are far better at filtering out noise in complex environments, search tasks inherently have less drain on mental resources and thus lower workload is experienced. The Tasks types: checking, detection, and response implementation all start with a search component to locate a specific I&C. While experts have similar efficiencies for locating controls, research has shown that the underlying process and techniques are not necessarily the same between experts (Jarodzka et al., 2010). Since patterns differ between experts, the “equal but different” principle cannot leverage training search heuristics as a method to reduce the knowledge gap between novices, experienced, and expert operators. Experts can leverage domain knowledge and years of experience to quickly filter out noise when searching in a complex environment. Information theory concepts provide researchers with a method for mitigating this gap in knowledge. Shannon equates information as a reduction in uncertainty (Cole, 1993). By systematically removing noise form the novice and experienced operator’s environment, the researcher is reducing the search space which provides information. The goal for the reduced search space is to provide a modified environment that has similar cognitive demands for novices and experienced operators as the full environment does for experts. For this research, the reduction in complexity was accomplished by proportionally removing ancillary I&Cs by their type such that the system maintains ecological validity.

Validating Novice Research Approach

The goal for this research is to validate the “equal but different” principle for using novices as a proxy for expert ROs when investigating levels and types of workload during NPP MCR operations. By stratifying participants in to three levels of experience based on the Dreyfus model: novice, experienced and expert, this research will provide a holistic view of the levels and types of workload observed in NPP MCR operations. Novices and experienced ROs will execute a modified OP following the “equal but different” principle that streamlines the instructions, simplifies the lexicon, and reduce visual complexity through information theory. On the contrary, experts will go through the same OP with the full lexicon for instructions and the full visual complexity of unmodified I&C panels. This will allow a comparison between the novice and expert group as well as a comparison between experienced and expert groups.

The aim for this research is to determine if certain workload measures are sufficiently effective across skill level at assessing types of workload experienced during NPP MCR task types. In addition to types of workload experienced, this research aims to measure trends in terms of workload level across the experience continuum form novice to expert and from experienced to expert ROs.

Since ROs work as a team to complete an OP, different roles within the team perform slightly different tasking. Therefore, comparisons between experience levels will only be made with participants that completed identical instructions within a task type. RO1 role will be performed by novice and expert participants and RO2 role will be performed by experienced and expert participants. That means analysis will not contain comparisons between novice and experienced groups. In addition to the fact that novices and experienced participants did not perform identical actions within each task type, directly comparing these two groups does not assist in answering the research questions in this dissertation. All comparisons between experience level will be made between novice and experts and between experienced and experts participants.

METHODOLOGY

Participants

The present study used participants from three discrete populations based on levels of experience in operating within a NPP MCR. The three populations were novice, experienced, and expert. The novice participant group for this study consisted of undergraduate and graduate student volunteers from the University of Central Florida (UCF). This group had no prior experience operating within a NPP MCR. The experienced population group consisted of extensively trained staff from the NRC HPTF at UCF. While the experienced group had no operational experience within a NPP MCR, they went through extensive training in a Generic Pressurized Water Reactor (GPWR) MCR Simulator. Lastly, the expert population group consisted of former ROs with operational experience in NPP MCR operations.

Novice participants (38 males, 31 females, *M=20.12, SD=2.51*) were recruited using UCF's psychology research participation system - SONA. In order for novice participants to participate in this experiment, they were required to have normal or corrected-to-normal vision and not have color vision deficiency. They were also required to abstain from ingesting nicotine for at least two hours prior to the experiment and abstain from alcohol and/or sedative medications for twenty-four hours prior to the experiment. Novices were compensated for their participation in the experiment through the form of research participation credits for their psychology class.

Experienced participants (2 females, ages 27 and 40) were volunteers selected from within the staff of the NRC HPTF at UCF. Each completed over 30 experimental sessions. The last six sessions (approximately 20% of their total sessions run) of each participant were used in the analysis. The present experiment classified an experienced RO as someone trained to perform specific rule-based and skill-based tasks commonly required of a RO in a NPP MCR, but does not have the full knowledge of or experience in an operational NPP MCR. Experienced participants were selected through an interview process with the Principle Investigator (PI) to determine if the volunteer was able to meet the physical and mental demands required for the duration of the study. The experienced participants were subject to the similar requirements as the novice group. They were required to have normal or corrected-to normal vision and not have color vision deficiency. In addition, experienced participants were required to abstain from ingesting nicotine for at least two hours prior to experimentation as well as abstain from alcohol and/or sedative medications for twenty-four hours prior to the experiment. Experienced participants were compensated for their participation in the present experiment through their hourly wage.

Expert participants (RO1: 9 males, 0 females, *M=48, SD=11.97; RO2:* 5 males, 4 females, *M=43.89, SD=9.36*) consisted of former ROs that have operational experience working in a MCR from the commercial power generation and/or military power generation domains. These ROs consisted of former pressurized water reactor (PWR) operators and/or boiling water reactor (BWR) operators currently employed by the NRC. Expert participants were compensated for their participation through the NRC.

Training

Novice participants

Novice participants went through a 2-hour training session prior to experimentation. This training was conducted using instructional materials delivered through PowerPoint and a GPWR NPP MCR simulator. Training consisted of 3-way communication, navigation within the simulated control room environment, as well as instruction on how to properly interact with I&Cs within the simulator. Novice participants were required to achieve an 80% level of proficiency on their evaluations for each communication, navigation, and execution on I&Cs in order to participate in the experiment. Failure to reach the 80% proficiency threshold resulted in the participant being dismissed.

Experienced participants

Experienced participants performed the role of RO2 and operated in a team with a RO1 and SRO. For all experimental sessions involving experienced participants, RO1 duties were performed by a novice, and SRO duties were performed by a researcher. Experienced participants repeated experimental sessions multiple times. To reduce the impact of team member interaction on performance between sessions, experienced participants were paired with the same researcher for the duration of the study.

RO2 experienced participants were trained to a significantly higher level of proficiency than the novice participants were for this study. Due to the extensive training and experimentation schedule demands only two participants were able to meet the qualifications of an experienced operator. Experienced operators underwent seven separate training sessions that lasted approximately seven and a half hours in total. This training consisted of in-depth instructional materials (e.g., PowerPoint slides), as well as time in the GPWR NPP MCR simulator. Time in the simulator was used to rehearse tasks similar to the operating procedures designed for this experiment.

In order to ensure a consistent baseline of knowledge for the experienced ROs, a comprehensive training curriculum for the specific tasking required in the experiment was developed. The curriculum consisted of two training manuals, a PowerPoint presentation, and multiple training evaluations. The first training manual covered materials on operating in the role of a RO in a PWR NPP MCR. The training manual was covered in three phases.

Phase 1 covered the basic structure of a GPWR NPP, expectations of the experienced participant during the experiment, and the roles and responsibilities of the individuals participating in the experiment (i.e., RO1, RO2, and SRO). Experienced participants always operated in the role of RO2 during experimentation. Phase 1 also covered the components of three-way communication along with scripts for the experienced participants to practice.

Phase 2 of the training manual covered the details of the GPWR NPP MCR simulator. In this section, experienced RO participants learned the layout and structure of the MCR I&C panels. Experienced participants learned how to identify, locate, and interact with light boxes, status boxes, gauges, and valves. Phase 2 concluded by providing the experienced participants with simulator time where they practiced searching for I&Cs on the MCR panels.

Phase 3 covered the experienced participants' specific responsibilities as RO2 for performing the modified OP on the simulator. The experienced participant practiced performing the modified OP for the three task types: checking, detection, and response implementation. After the experienced participant completed training for each of the task types, they completed a section evaluation. A section evaluation consisted of similar tasks just trained in which the experienced participant demonstrated proficiency with the simulator on the specific task. The experienced participant was required achieve 80% or higher or they were dismissed. This training lasted about two hours for each experienced participant. The phase 3 section evaluation was completed two separate times by each experienced participant.

In addition to the RO functionality training discussed in the training manual above, a second training manual with additional training sessions was also developed. This manual covered appropriate verbiage to use during experimentation, performance expectations, experimentation scheduling, and other miscellaneous participation requirements. A companion PowerPoint presentation with a verbal training script was presented to the experienced participants. This training was also divided into multiple sections. Each section had a corresponding evaluation. The assessment included learning objective based questions, assessing the knowledge and information gained from the material presented. For an experienced participant to continue, they were required to score a 90% or better. This training session lasted around an hour for each participant. This training was completed twice by each experienced participant before they moved on to the mock experimentation.

Mock experimentation sessions lasted around a half hour. Experienced participants perform the modified OP used in the experiment as a RO2 with a paired SRO. Experienced participants were evaluated during the mock sessions by the paired SRO. The SRO provided feedback to the participant when needed. Experienced participants completed three separate mock experimental sessions with their paired SRO before they participated in recorded experimental sessions.

In all, experienced participants completed a minimum of seven training sessions lasting around seven and a half hours. These training sessions were rigorous and time consuming, but provided a consistent baseline as the starting point for the knowledge required for an experienced RO.

Expert participants

Expert participants went through a 2-hour training session. This training was conducted using PowerPoint and the GPWR NPP MCR simulator used in the experimental sessions. Training was required to ensure expert participants were given an opportunity to become familiarized with the specific configuration of the GPWR NPP MCR simulator. This training was required because each operational NPP MCR has slightly different operating procedures. Training covered 3-way communication with the specific lexicon for the GPWR NPP MCR simulator, as well as how to navigate and correctly manipulate I&Cs within the simulator.

Experimental Environment

Experimental sessions were conducted in a laboratory room setup as a mock MCR. A GPWR NPP MCR simulator was configured for a crew of three operators. Crews consisted of two ROs and a SRO. The names RO1, RO2, and SRO, were used to refer to the crew members during the experiment. The SRO role was played by the researcher and the participants operate in the role of the two ROs. Data collection was conducted at two locations. The novice and experienced participant experimental sessions were conducted in the NRC HPTF at the UCF (Figure 2). Expert participant experimental sessions were conducted in the NRC HPTF at the NRC headquarters in Rockville, MD (Figure 3). The laboratory layout and configuration for the expert participants was setup similar to the laboratory layout and configuration used for the novice and experienced participants.

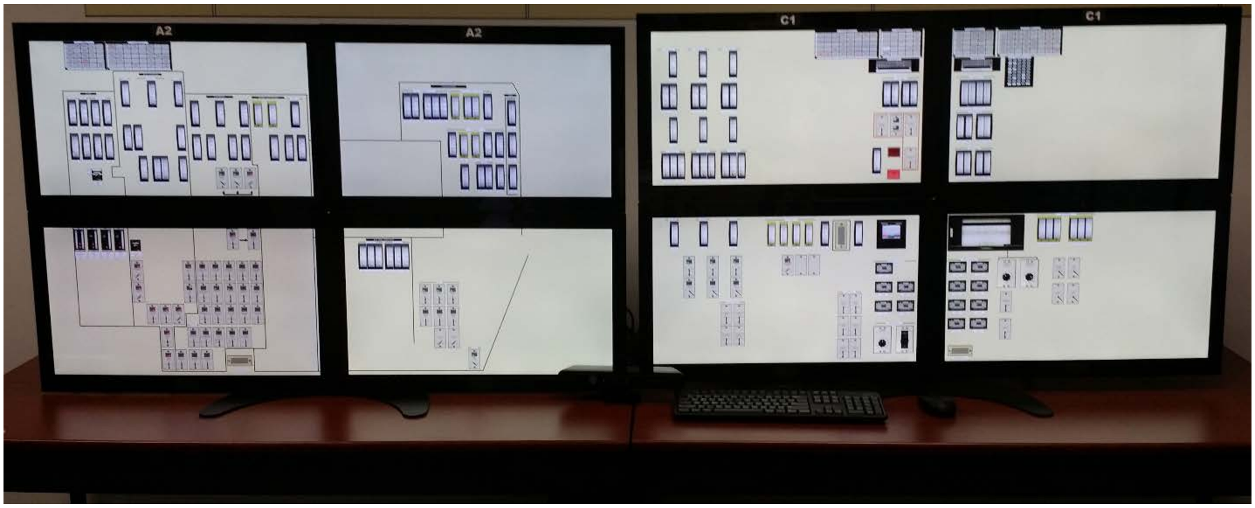

Figure 2: Simulator with the RO2 panels setup at the HPTF at UCF

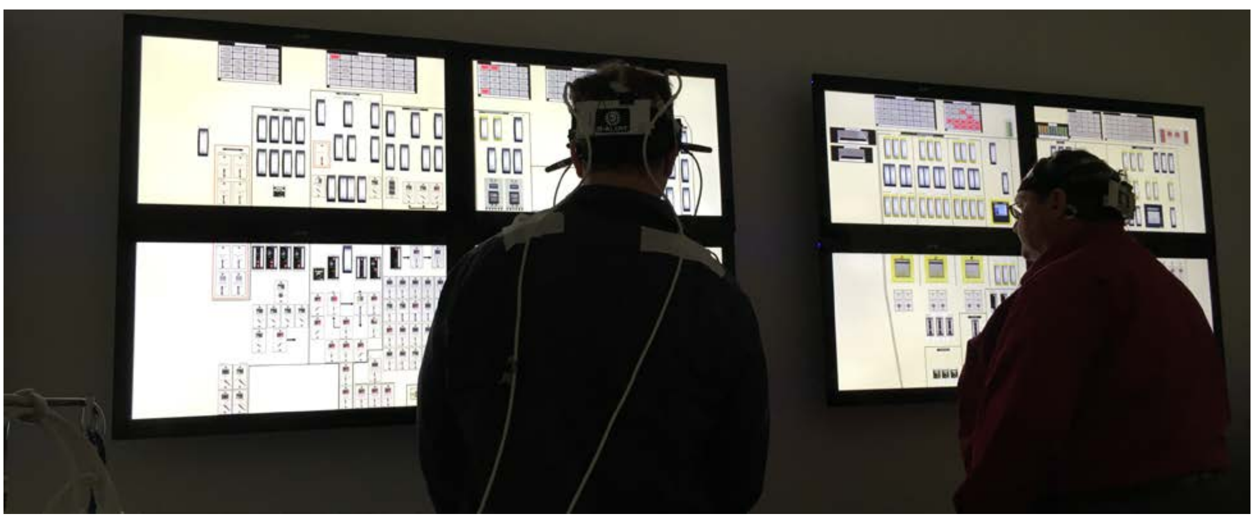

Figure 3: Simulator at the HPTF at NRC headquarters (RO1 is on the left, RO2 is on the right)

Simulator Hardware

For both experimental locations, the simulator hardware consisted of four identical workstation computers networked together on a gigabit network backbone. The hardware specs for the workstation computers were: Xeon X5650 6 core processor, Matrox M9188 graphics card (UCF location) or GeForce GTX 970 (NRC location). In addition to the workstations, Microsoft Kinects were used to record communication events that occured during the experimental conditions. All of the simulator's software ran on the Microsoft Windows 7 64bit operating system. The software ran on java 1.7 SE Runtime Environment.

Simulator Software

There are several applications that together made up the software component for GPWR NPP MCR simulator: JDesigner™, GPWR™, and EPIC. JDesigner™ was the interface design tool used to construct virtual panels for the NPP MCR. GPWR™ was the full-scope model of a generic NPP that links the panel's I&Cs to the physics of the PWR. Experimental Platform for Instrumentation and Controls (EPIC) software was an in-house developed software package that mimicked the user interface panels from JDesigner™; however, it controlled the I&C states on the panel from defined scripts rather than from reactor physics models. EPIC software provided a repeatable experience between participants.

Displays

Each participant operated on two control room wall panels (Figure 4). Panels consisted of four 27^{in} monitors arranged two high by two wide. Each 27^{in} monitor had a resolution of 2560 pixels by 1440 pixels tall.

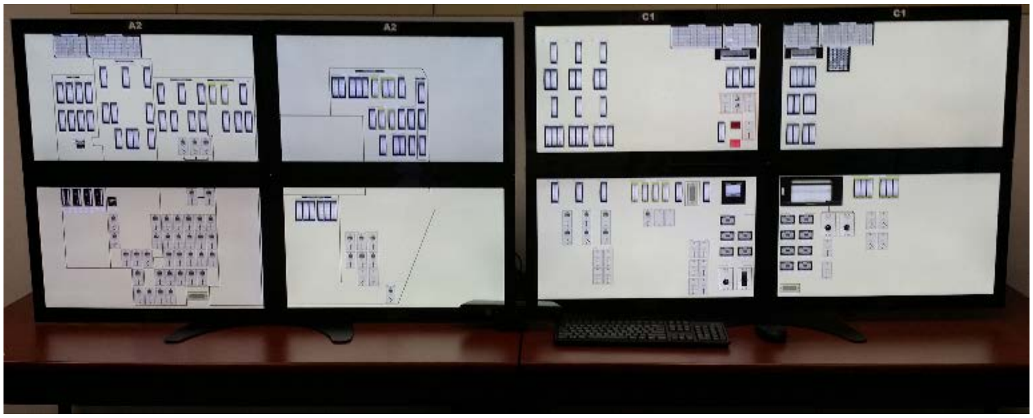

Figure 4: Simulator with two control room wall panels

Physiological Instruments

The Physiological sensor hardware that was used for this experiment was Advanced Brain Monitoring's B-Alert X10. The X10 has a 9-channel electroencephalogram (EEG) Bluetooth system that covered the mid-line and lateral EEG sites (Figure 5) of the participants. The X10 also has a tenth channel for recording electrocardiogram (ECG). The X10 has a sample rate of 256 Hz. Custom synchronizing software was utilized for capturing raw and filtered EEG and ECG waveform signals during experimentation. This software linked tagged events in the simulation to the recorded physiological signals.

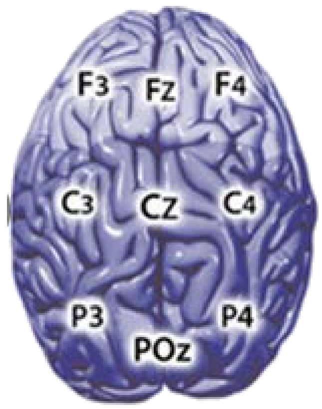

Figure 5: B-Alert EEG locations

Subjective measures

DUJO software was utilized in this experiment to present surveys to the participants in a digital form. This software tied the participants' subjective survey responses and physiological sensor data with the EPIC simulation through an application programming interface (API). This enabled marked simulation events to be compared to participant's physiological state with millisecond accuracy.

Scenario Initialization

The initial condition for the state of the NPP was a total loss of AC power to the 1A-SA and 1B-SB safety buses. EOP-EPP-001 is the EOP to follow in the event of a loss of AC power to the 1A-SA and 1B-SB safety buses (G. P. Systems, 2011). Modifications were made to the EOP procedures with Subject Matter Experts (SME) input to focus tasking on two panels per RO. For the novice and experienced participants, MCR panels had a reduced set of I&Cs as a way to reduce visual complexity. The modifications to the EOP along with justifications will become clearer in the next few paragraphs.

Reactor Operator Roles

Crews of three operators are required to complete the steps outlined in EOP-EPP-001. For the present experiment, the roles were called RO1, RO2 and SRO. Novice or expert participants performed the actions required for the crew role of RO1, respective of experiment location. RO2 role was performed by experienced or expert participants, respective of experiment location. The researcher performed the duties of the SRO role irrespective of experiment location. The SRO role was the leadership role that initiates and coordinates RO1 and RO2 tasking in order to accomplish the EOP correctly and efficiently. Experienced operators repeated the experiment multiple times in the role of RO2 at the UCF location. While several different researchers performed the role of SRO, the experienced RO2 was always paired with the same researcher throughout all experimental trials.

Complexity reduction

The actual steps for completing EOP-EPP-001 required ROs to interface with five different I&C panels (A2, B1, B2, C1, and D1). SMEs provided input to modify the EOP procedure to reduce complexity in a way that each RO only interfaced with two panels. RO1 interfaced with panels A2 and C1. While RO2 interfaced with panels A2 and B1. In addition to modifying the EOP to reduce panels, supplemental steps were added to the modified EOP as a way to maintain experimental control across the three task types.

A further reduction in system complexity was implemented for experimental sessions with novice and experienced participants only. To match complexity of the tasking environment to level of experience, the full I&C was used for the experts. This complexity reduction was implemented to normalize and reduce search space for locating I&Cs. Panels A2, B1 and C1's was normalized such that each panel contains an identical number of I&Cs. The normalization procedures maintained each panel's correct proportion of I&C types. I&C's on the panels in an MCR were broadly categorized in to five control types: light boxes, status boxes, gauges, valves, and miscellaneous other controls. The panel complexity reduction procedure was adopted from Mercado (2014) experimental design. The procedure that was used to normalize the panels was as follows. For each panel used in the EOP, the total number of all control types ware counted. The control count from the panel with the fewest number of controls was selected as the baseline for the remaining panels (i.e., in the present study, the baseline was 113 based on panel C1 having the fewest number of controls). Next, the normalizing factor for each panel was calculated using the formula: baseline control count from panel C1 and divided by number of controls on the current panel (i.e., $normalize\_factor = \left[\frac{I\&C\_count\_panel\_c1}{I\&C\_count\_current\_panel}\right]$). For each panel, the number of controls within each category was counted (e.g., light boxes, status boxes, gauges, and misc.). The reduced amount for each control category was calculated by multiplying the number of controls in the category times the normalizing factor. Lastly, the number of controls in each category was rounded such that the total number of controls in all categories for a panel was equal to the baseline number on controls on panel C1. Only controls on a panel that were not used in the modified EOP procedure were randomly selected for removal. See tables below for the calculated number of controls used for the A2 (Table 1) and B1 (Table 2) panels during the novice and experienced experimental sessions.

Table 1: A2 Panel Modification

Control Categories	# Controls on A2 Panel	Normalize Factor 57.36%	Reduced # of Controls	Number of Controls on modified panel
Light Boxes	4	* 0.5736	2.2944	2
Status Boxes	0	* 0.5736	0.0	0
Gauges	80	* 0.5736	45.88	46
Valves	108	* 0.5736	61.9488	62
Miscellaneous	5	* 0.5736	2.868	3
Total	197		113	113

Table 2: B1 Panel Modification

Control Categories	# Controls on B1 Panel	Normalize Factor 71.06%	Reduced # of Controls	Number of Controls on modified panel
Light Boxes	4	* 0. 7106	2.8427	3
Status Boxes	3	* 0. 7106	2.1320	2
Gauges	78	* 0. 7106	55.4340	55
Valves	41	* 0. 7106	29.1384	29
Miscellaneous	33	* 0. 7106	23.4528	24
Total	159		113	113

The majority of controls on I&C panels in an NPP MCR contain two identifiers, an abbreviated control name and an alphanumeric code. In an operational setting when executing an EOP, the ROs refer to controls using their full control name (e.g., Steam Generator Alpha Narrow Range Channel 1). For experts, the control's full name provides context useful for locating controls and abbreviations are understood. Since steam generators are collocated, if instructed to report the Steam Generator Alpha Narrow Range Channel 1 level, the RO can use I&C panel layout domain knowledge for narrowing their search space when locating the specific control. For novices and experienced participants to execute the EOP using the controls full name would require additional domain knowledge training. Therefore, for the novice and experienced participants, the EOP was modified to refer to the controls by their alphanumeric code rather than the controls full name (Figure 6). For the expert experimental sessions, controls were referred to by the controls full name along with the alphanumeric code.

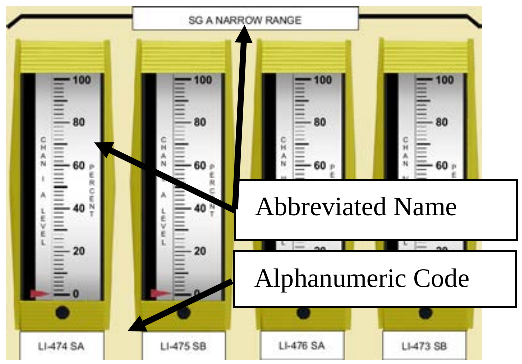

Figure 6: Gauge Naming Example. Alphanumeric code name was used rather than the full gauge name

The length of some of the alphanumeric codes exceed the recommended seven plus or minus two characters defined in Miller’s Law (Miller, 1956). It is widely accepted in the academic community that an average human can hold around seven objects in working memory. To mitigate working memory confound for novice and experienced participants, the researcher recoded all alphanumeric codes larger than seven so that they contained seven or fewer characters (e.g., FI-2050A1 SA was changed to FI-25A1) (Figure 7). Because the alphanumeric codes have meaning for ROs, recoding them for experts would have obfuscated instructions by having a mismatched control name and alphanumeric name. Therefore, recoding only occurred for the novice and experienced sessions.

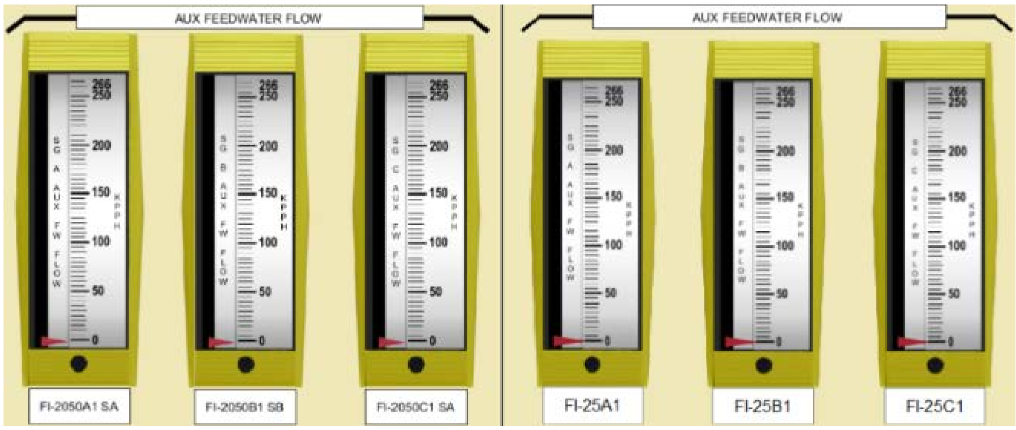

Figure 7: Example gauges with renamed alphanumeric codes. Left three gauges contain full alphanumeric codes. The right three gauges are identical to the ones on the right with the recoded alphanumeric code of seven or less characters

EPIC and JDesigner/GPWR

In this experiment, participants interacted with the virtual MCR panels through EPIC software. EPIC leveraged design and functionality from JDesigner™ and GPWR™, respectively. Each operational NPP has a customized MCR layout. Because each MCR is uniquely configured, JDesigner™ was needed for designing the virtual MCR panels used in the present study. GPWR™ is a training simulator used to teach fundamentals of a PWR ("Training Applications | GSE Systems," 2015) and supports representing the physics of a NPP with ANS-3.5 (G. Systems, 2015) compliant models. The GPWR™ software was used during the experimental design phase to develop and evaluate task procedures. Specifically, it was used for timing gauge value changes, setting all gauges to proper initial values, and for evaluating the accuracy of the modified EOP task procedures. Because GPWR™ models are based on NPP ANS-3.5 physics models, the software is designed using pseudo random processes and not capable of providing identically repeatable events across experimental trials. EPIC utilized the panel designs from JDesigner™ and gauge change timing from GPWR™. EPIC software was specifically designed as a tool for experimentation and provides millisecond accuracy for logging participant interactions.

Experimental Design

A mixed 3 (*Task Type*: checking, detection, response implementation) × 2 (*Experience Level*: novice, expert) Analysis of Variance (ANOVA) design was employed in the present experiment with repeated measures on task type for the role of RO1. Identical analysis was repeated for experienced and experts in the RO2 role. A mixed 3 (*Task Type*: checking, detection, response implementation) × 2 (*Experience Level*: experienced, expert) for the role of RO2 was employed with repeated measures on task type. In both analyses, task type was partially counterbalanced to maintain external validity. NPP MCR procedures require performing a checking of I&C's state before executing a response implementation on an I&C. Therefore, task yoking was observed to maintain external validity in each session. Each session was randomly assigned one of three presentation orders (Table 3).

Table 3: Partial Counterbalanced Task Ordering

Presentation Order	**Task Type**		
	Condition 1	Condition 2	Condition 3
1	Checking	Response Implementation	Detection
2	Checking	Detection	Response Implementation
3	Detection	Checking	Response Implementation

In each experimental session a total of twelve tasks were performed (four checking, four detection, and four response implementation). To ensure ecological validity with the EOP, the four steps within each task type were always performed sequentially in the identical order across sessions.

Independent Variables

The independent variables for the experimental design ware task type (i.e., checking, detection, and response implementation), role (i.e., RO1 and RO2), and experience level (i.e., novice, experienced, and expert).

Task Type

For the present experiment, three conditions were used to investigate task type. The three conditions consisted of checking, detection, and response implementation tasks. For each condition, only one task type was used and consisted of performing the operations required for the task type a total of four separate times, each with a different I&C in accordance with the modified EOP. All tasks were initiated by the SRO through a three-way communication instruction.

The checking task type required the participant to inspect the state of a light box or valve to confirm or deny that it is in the desired state as defined in the EOP. Participants were required to use three-way communication with the SRO to verify they understood the checking instruction given to them by the SRO. Then they searched the panels to locate the I&C defined in the instruction. Once located, they were required to touch the I&C, signaling they located it and have determined its current state. Lastly, to complete the task, participants initiated a three-way communication information event notifying the SRO the current state of the light box or valve.

The detection task type required participants to continuously monitor a gauge's level and report when it reaches a threshold. The task started when the SRO issued a three-way communication instruction event, identifying to the participant the gauge to monitor and the level at which to report back once it crosses. Participants were then required to use three-way communication with the SRO to verify they understand the detection instruction given to them by the SRO. Then they searched the I&C panels to locate the gauge defined in the instruction. Once located, they touched the gauge, signaling they have located it. Touching the gauge triggered a five-minute script that changed the gauge's level. Twelve random times a minute, the gauge value fluctuates in noticeable discrete changes. A total of fifty-nine changes occurred over the five-minute script before the gauge value crossed over the reporting threshold for the sixtieth change event. The participant was required to signal they noticed each of the sixty changes in level by clicking the gauge label button directly below the gauge immediately after noticing a change in its value. After the gauge crossed the threshold value, participants are required to initiate a three-way communication information event to notify the SRO the gauge had crossed the threshold value.

The response implementation task type required participants to take an action that will alter the state of the NPP. This task required fine motor skills which required participants to rotate a virtual switch on the touch panel. Similar to checking and detection, participants used three-way communication with the SRO to verify they understood the response implementation instruction (i.e., which valve to manipulate and weather to open or shut the valve). Then they located the correct valve on the I&C panel in accordance with the SRO instruction. Once located, participants were required to touch the valve to signal they had located it. Once touched, they manipulated the valve to the left or right to open or shut the control in accordance with the SRO instruction. Upon completing the open or shut manipulation, the participant initiated a three-way communication information event to notify the SRO the valve had been open or shut.

Dependent Variables

Subjective Measures

Instantaneous Self-Assessment (ISA) was used to measure overall perceived workload online during the task type. The ISA is a one question survey with a five-point Likert scale (Tallersall & Foord, 1996) where participants provide their perceived workload at the instant the question is asked. Prior to experimentation, participants were given definitions and instructions on how to complete the ISA survey. Halfway through a condition (i.e., participant has completed two of the four instructions initiated by the SRO) an audio prompt was triggered by a script. The audio prompt said “please rate your workload”. Participants respond orally with a number 1 through 5.

The NASA-Task Load indeX (TLX) developed by Hart and Staveland (Hart & Staveland, 1988) was used to evaluate subjective workload experienced by the ROs using a multi-dimensional scale with subscales. The subscales used were mental demand, physical demand, temporal demand, effort, frustration, and performance. The TLX questionnaire asked the participant to rate each subscale on a 0 to 100-point scale with a five-point increment resolution. Participants were given a copy of the scale with subscale definitions to reference during the questionnaire. A global workload measure was calculated by averaging the six subscales. The TLX survey was presented to the participants at the end of each condition using a computer version of the survey.

Performance

The checking task performance consisted of correctly identifying the target control per the SROs instruction. The simulator logged all touches to determine if the correct I&C was identified or if the participant selected an incorrect I&C. In addition to the simulator logging all touch events, audio recordings were collected to evaluate verbal verification of the correct light box or valve state.

The detection task performance consisted of correctly identifying the I&C as in the checking task. In addition to the correct identification, the detection task type had a gauge change detection component. The simulator logged participant's hits, misses, and false alarms on the gauge change portion of the detection task. Audio recordings were collected to evaluate verbal verification by the participant when the gauge's level reached or crossed the specified level.

The response implementation task performance consisted of correctly identifying the I&C identical to both the checking and detection tasks. In addition to correct identification, the response implementation task had a valve manipulation component. The simulator logged correct and incorrect manipulations made by the participant. Audio recordings were collected to evaluate verbal verification by the participant that the valve has been manipulated correctly (opened or closed) per the SROs instruction.

Communication Performance was evaluated against a pre-defined grading rubric. Three-way communication rules was used to construct the rubric. Three-way communication consists of two parties in a back and forth manor clearly stating an instruction or informing the state of an I&C. A typical three-way communication consists of the first party initiating an instruction or information to a second party. The second party gives acknowledgement of the instruction or information by repeating the specific details they understood. Lastly, the first party closes the loop by confirming or denying the acknowledgement. Within a three-way communication, either party may provide clarification, request location assistance or request a repeat of the last statement. Audio recording for each condition were manually transcribed and encoded in to the three-way communication events. The performance metrics were calculated as follows: number of instruction events completed per task, number of instructions that were required to be repeated by the participant, number of instructions that required clarification by the participant, number of times I&C location help was requested by the participant and percent correctly executed three-way communications. The number of instruction events completed per task were defined as the three-way communication events that the RO successfully completed. The number of instructions that were required to be repeated is a count of the number of times the RO requested the SROs instruction to be repeated. The number of instructions that required clarification is a count of the number of times the participant incorrectly acknowledged the SRO instruction (e.g., repeated back an incorrect I&C name or an incorrect gauge threshold value). The number of times I&C location help was requested by the participant is a count of the number of times the RO was having trouble locating the desired I&C and requested help in locating the control. Percent correct was scored by comparing the participant's three-way communication events against the ideal responses in the rubric.

Physiological Measures

Electroencephalogram (EEG)

This experiment followed the international 10-20 system for electrode placement (Klem, Lüders, Jasper, & Elger, 1999). The B-Alert X10 system used in this experiment had a sampling rate of 256 Hz and captured signals from Fz, F3, F4, Cz, C3, C4, POz, P3, and P4 regions (Figure 8). A reference electrode was placed on the mastoid bone. Power-spectral-density (PSD) analysis was conducted on the filtered EEG signals to investigate frequency responses by bands [theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz)] (Thakor & Tong, 2004). In addition to PSD analysis at each electrode site, PSD analysis was also calculated for groups of sites. The sites analyzed were left hemisphere, right hemisphere, frontal, parietal, and occipital lobes.

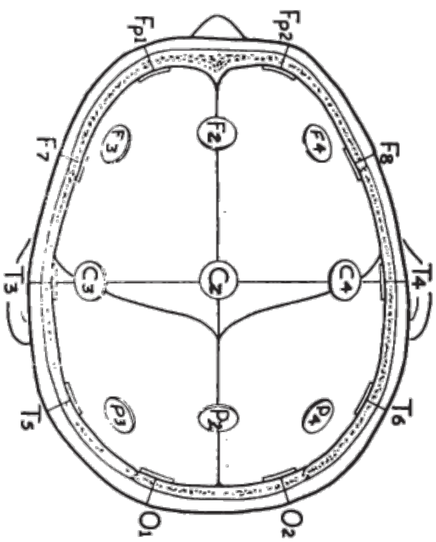

Figure 8: Electrode placement for EEG

Electrocardiogram (ECG)

The B-Alert X10 was the system that was used in this experiment to collect the ECG waveform. This system used two single lead electrodes. The anode is placed on the participant's right clavicle. The cathode is placed on the lowest right rib (Figure 9). Three measures were derived from the ECG waveform, heart rate (HR), interbeat interval (IBI), and heart rate variability (HRV). The "So and Chan" QRS detection method was used to calculate R-Peaks (So & Chan, 1997) to compute HR, IBI and HRV measures.

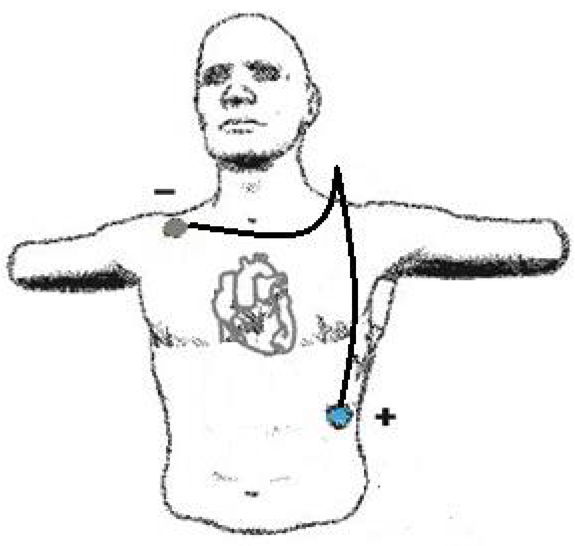

Figure 9: Electrode placement for ECG

Procedures

Novice Participants

Novice participants received an informed consent. Upon reading and signing the consent form, participants received the Ishihara Color Vision Test, restrictions checklist and demographics survey to complete. Novice participants then went through training with the researcher lasting around two hours. The NPP MCR functional areas that the training covered were: procedures and protocols for operating in an NPP MCR, three-way communication, interacting with the EPIC simulated MCR, and how to read/interpret status and light indicators. The researcher guided the participants through training in each functional area in isolation. Participants then reinforced the skills learned in training through a unified practice session incorporating all functional areas. The researcher administered a proficiency test at the conclusion of each functional training area and again during the practice session. In order to continue with the study, novice participants were required to score 80% or higher on all proficiency tests. Novice participants were given a five-minute break after completing the training.

The B-Alert EEG and ECG sensors were placed on the novice participants and a five-minute wakeful rest baseline was conducted. After the baseline, the three experimental conditions were performed sequentially (in the presentation order determined for the participant). Novice participants perform the role of RO1. The SRO used three-way communication during each condition to initiate each of the four tasks designated for RO1 within the task types. An audible ISA prompt was played through computer speakers half way through each task type (i.e. immediately following the completion of step two of four). Upon completion of each task type, the NASA TLX questionnaire was administered. Lastly, upon completion of the three task types and questionnaires, the B-Alert sensors were removed and the participant was dismissed.

Experienced Participants

A researcher explained the purpose of the study and provided experienced participants an informed consent prior to preparing them for the experienced RO role. Once the participant had read, understood, and signed the informed consent, they were issued the Ishihara Color Vision

Test (Ishihara, 2010). A passing score of 90% or higher was required to continue participating in the study. Participants then completed a demographics questionnaire and went through training. Training covered everything required to perform the role of RO2 in this specific experiment at the proficiency level of an experienced operator. The researcher administered evaluations following each functional area trained and experienced participants were required to score 90% or better to continue with the experiment.

Upon completion of training and initial surveys, experienced RO participants were eligible to participate with their paired SRO researcher in experimental sessions. Experienced RO participants were paired with novice RO participants. Novice participants always performed the RO1 duties and experienced participants always performed RO2 duties. Each experimental session started by placing the B-Alert EEG and ECG sensors on the experienced participant. The experienced participant then completed a short practice session performing the duties of RO2 along with the crew (RO1 and SRO). At the conclusion of the practice, the experienced participant completed a five-minute wakeful resting baseline. After the baseline, the three experimental conditions were performed sequentially (in the presentation order determined for the participant). During the experimental conditions, experienced participants performed the duties required of the role of RO2. The SRO used three-way communication during each condition to initiate the four tasks designated for RO2. An audible ISA prompt was played through computer speakers half way through each task type (i.e. immediately following the completion of step two of four). Upon completion of each task type, the NASA TLX questionnaire was administered. Lastly, upon completion of the three task types and questionnaires, the B-Alert sensors were removed and the participant dismissed.

Expert Participants

The expert participants were provided an informed consent. Upon reading the informed consent, a demographics survey was administered. Since each NPP has slightly different operating procedures, training was conducted on the specifics of this simulated PWR NPP MCR to insure consistent expectations across the expert participants. Training for the expert participants was conducted by the same researcher using PowerPoint and the EPIC simulator. The training lasted around two hours for each RO1 and RO2 team. At the conclusion of the training, expert participants were given a five-minute break.

The B-Alert EEG and ECG sensors were placed on the expert participants and a five-minute wakeful rest baseline was conducted. After the baseline, the three experimental conditions were performed sequentially (in the random presentation order pre-determined for the participant). Expert participants performed in either the role of RO1 or RO2. The SRO used three-way communication during each condition to initiate the four tasks designated for RO1 or RO2 within each task type. An audible ISA prompt was played through computer speakers half way through each task type (i.e. immediately following the completion of step two of four). Upon completion of each task type, the NASA TLX questionnaire was administered. Lastly, upon completion of the three task types and questionnaires, the B-Alert sensors were removed and the participant dismissed.

RESULTS

Analysis

All analyses were computed using SPSS v24. Unless otherwise stated, rejection of the null hypotheses is set at 0.05 ($\alpha < 0.05$). Data analysis was divided in to two sections, (i) novice and expert participants that performed in the RO1 role as well as (ii) experienced and expert participants that performed in the RO2 role. The two experienced participants each completed around 30 experimental sessions; however, only their last six each were included in the analysis. This is because the last experimental sessions is where they have the most experience. The last six sessions (approximately 20% of their sessions) was choosen to reduce the impact of an outlier session impacting the analyses; however, still it include their most experienced sessions. Within each section, ANOVAs were used to assess workload across task types and experience level. When Mauchly's test of sphericity was violated ($\alpha < 0.05$), Greenhouse-Geisser corrections were reported. Bonferroni corrections were used in multiple comparison analysis to compensate for the Type 1 error rate increase. Brown-Forsythe statistic is reported when homogeneity of variance is violated for one-way ANOVAs. Effect sizes, sample means, and standard deviations were reported when appropriate.

RO1 Analysis (Novice and Expert)

Subjective Measures

Instantaneous Self-Assessment

A mixed 3 (*Task Type*: checking, detection, and response implementation) × 2 (*Experience Level*: novice and expert) ANOVA was conducted. Task type was a repeated measure factor while experience level was a between-participants factor. The ANOVA was used to determine if task type and experience level have significant effects on online-subjective workload (i.e., reported as the task was being performed), and to determine if the pattern of workload differences found across the tasks differed between novices and experts. For ISA, no main effect was found for task type for RO1 participants, ($p = .43$). A main effect for experience level was found for ISA, $F(1, 76) = 6.25, p = .02, {\eta_p}^2 = .08$, such that novice participants ($M = 2.19, SD = 0.50$) reported higher ISA compared to expert participants ($M = 1.74, SD = 0.52$).

There was a significant interaction effect between the task type and experience level of the participant, $F(1.64, 124.62) = 3.48, p = .04, {\eta_p}^2 = .04$ (Figure 10). Checking, $F(1, 77) = 11.54, p < .01, d = -1.21$, elicited higher ISA for novice participants ($M = 2.41, SD = .60$) than for expert participants ($M = 1.67, SD = 0.71$). Detection, ($p = 1.0$) , did not elicit different ISA between novice ($M = 2.0, SD = 0.92$) and expert participants ($M = 2.0, SD = 0.71$). Response implementation, $F(1,77) = 10.28, p < .01, d = -1.13$, elicited higher ISA for novice participants ($M = 2.16, SD = 0.53$) than for expert participants ($M = 1.56, SD = 0.53$).

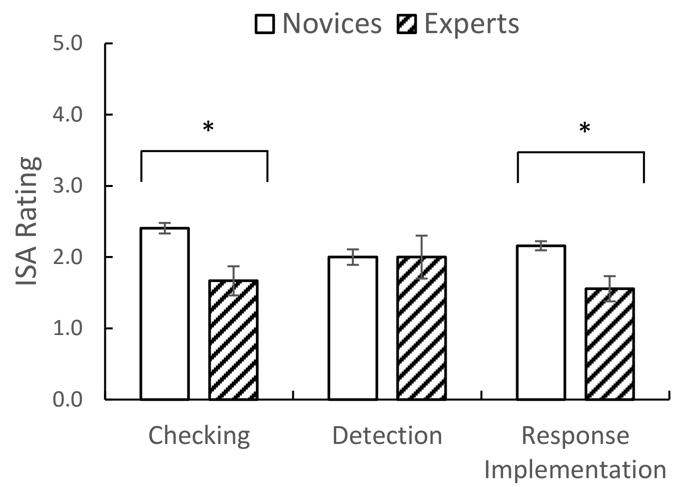

Figure 10: RO1 mean ISA by task and experience level. Error bars represent standard error. Asterisk represents statistical significance.

NASA Task Load Index (TLX)

A mixed 3 (*Task Type*: checking, detection, and response implementation) × 2 (*Experience Level*: novice and expert) ANOVA was conducted for NASA TLX global workload as well as each of the six subscales (mental demand, physical demand, temporal demand, performance, effort, and frustration). Task type was a repeated-measure factors, and experience level was a between-participants variable. The ANOVAs were used to determine if there was a significant workload difference between task types, and to see if there were differences in the ratings between the two levels of experience.

Task type had a significant effect on TLX global workload and several of the subscales that persisted across experience level. A main effect was found for task type for the TLX global workload, $F(2, 152) = 7.59, p < .01, {\eta_p}^2 = .09$, such that detection ($M = 32.67, SD = 19.84$) was higher than checking ($M = 27.68, SD = 17.18$), and response implementation ($M = 28.57, SD = 18.69$) did not differ from either checking or detection. A main effect for task type was found for physical demand, $F(1.73, 131.64) = 15.74, p < .01, {\eta_p}^2 = .17$, such that detection ($M = 31.03, SD = 3.53$) was higher that both checking ($M = 20.51, SD = 14.80$) and response implementation ($M = 24.23, SD = 3.27$). A main effect for task type was also found for frustration, $F(1.62, 123.26) = 13.60, p < .01, {\eta_p}^2 = .15$, such that detection ($M = 48.59, SD = 29.44$) was higher that both checking ($M = 24.94, SD = 24.46$) and response implementation ($M = 25.38, SD = 23.11$).

Experience level had an overall effect on the TLX global workload as well as several of the subscales that persisted across the various task types. A main effect was found for experience level on global workload, $F(1, 76) = 4.06, p = .05, {\eta_p}^2 = .05$, (Figure 11) such that novice participants ($M = 30.95, SD = 16.19$) reported higher global workload than expert participants ($M = 19.57, SD = 13.61$). A main effect for experience level was found for temporal demand, $F(1, 76) = 5.65, p = .02, {\eta_p}^2 = .07$, such that novice participants ($M = 32.34, SD = 20.85$) reported higher temporal demand compared to expert participants ($M = 15.37, SD = 12.82$).

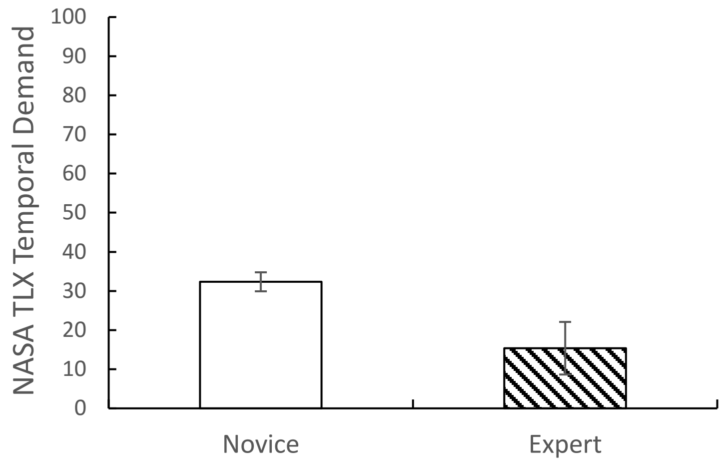

Figure 11: RO1 NASA TLX mean temporal demand by experience level. Error bars represent standard error.

A main effect for experience level was found for performance, $F(1, 76) = 6.88, p = .01, {\eta_p}^2 = .08$, (Figure 12) such that novice participants ($M = 31.14, SD = 20.63$) reported higher performance compared to expert participants ($M = 12.41, SD = 15.50$).

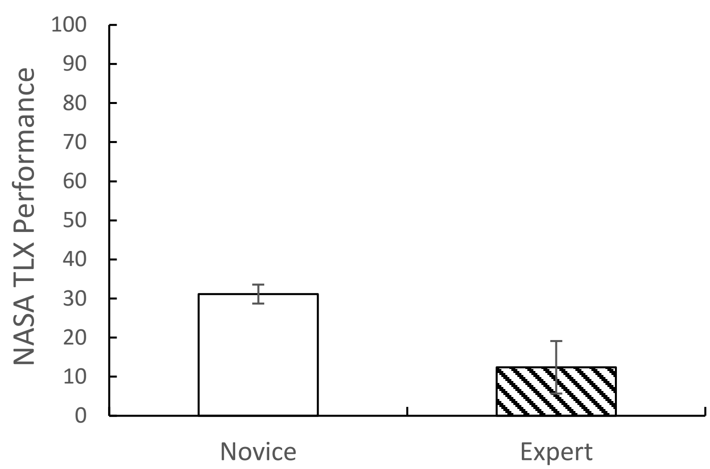

Figure 12: RO1 NASA TLX mean performance by experience level. Error bars represent standard error.

Interaction effect between task type and experience level was found for TLX global workload as well as several of the subscales. An interaction effect was found between task type and experience level for the global workload, $F(2, 152) = 3.44, p = .03, {\eta_p}^2 = .04$ (Figure 13). Checking, $F(1, 16.98) = 27.13, p < .01, d = -1.15$, elicited higher global workload for novice participants ($M = 29.83, SD = 16.87$) than for expert participants ($M = 11.20, SD = 8.83$). Detection, ($p = .55$), did not elicit different global workload for novice participants ($M = 33.16, SD = 19.74$) than for expert participants ($M = 28.89, SD = 21.43$). Response implementation, ($p = .09$), did not elicit different global workload for novice participants ($M = 29.86, SD = 18.79$) and expert participants ($M = 18.61, SD = 15.27$).

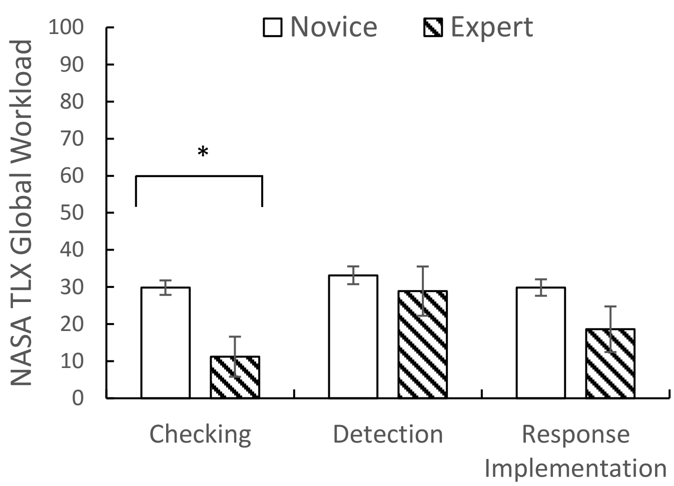

Figure 13: RO1 NASA TLX mean global workload by task type and experience level. Error bars represent standard error. Asterisk represents statistical significance.

An interaction effect was found between task type and experience level for physical demand, $F(2, 152) = 3.44, p = .03, {\eta_p}^2 = .04$, (Figure 14). Checking, $F(1, 29.87) = 17.02, p < .01, d = -0.728$, elicited higher physical demand for novice participants ($M =$

$29.83, SD = 16.87$) than for expert participants ($M = 11.20, SD = 8.83$). Detection, ($p = .55$), did not elicit different physical demand for novice participants ($M = 33.16, SD = 19.74$) and expert participants ($M = 28.89, SD = 21.43$). Response implementation, $F(1, 23.00) = 14.71, p < .01, d = -0.74$, elicited higher physical demand for novice participants ($M = 29.86, SD = 18.80$) than for expert participants ($M = 18.61, SD = 15.27$).

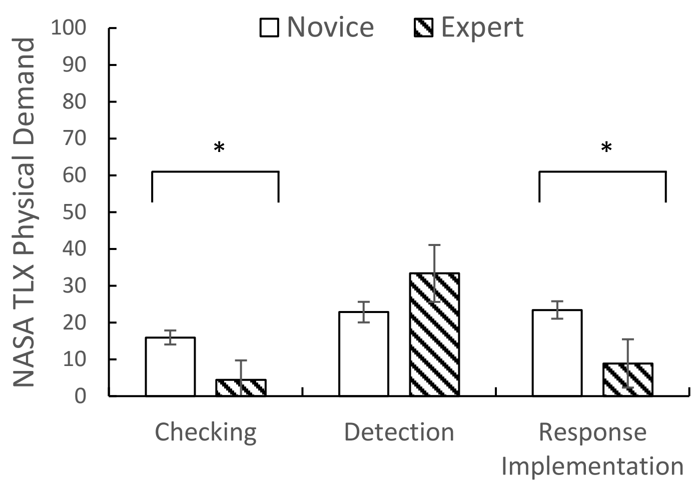

Figure 14: RO1 NASA TLX mean physical demand by task type and experience level. Error bars represent standard error. Asterisk represents statistical significance.

An interaction effect was found between task type and experience level for effort, $F(2, 152) = 3.60, p = .03, {\eta_p}^2 = .05$, (Figure 15). Checking, $F(1{,}20.38) = 18.44, p < .01, d = -0.87$, elicited higher effort for novice participants ($M = 29.28, SD = 21.88$) than for expert participants ($M = 11.11, SD = 9.93$). Detection, ($p = .85$), did not elicit different effort for novice participants ($M = 27.32, SD = 23.48$) and expert participants ($M = 28.89, SD = 22.75$). Response implementation, ($p = .22$), did not elicit different effort for novice participants ($M = 30.07, SD = 22.34$) and expert participants ($M = 20.56, SD = 15.70$).

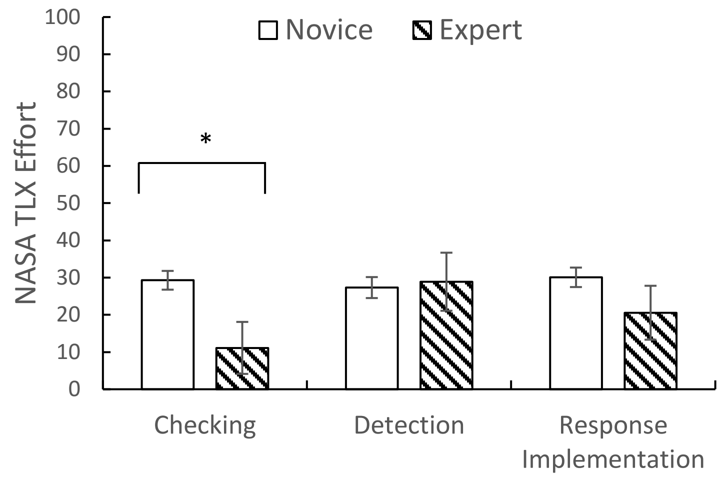

Figure 15: RO1 NASA TLX mean effort by task type and experience level. Error bars represent standard error. Asterisk represents statistical significance.

Performance Measures

Communication Reporting

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. Four mixed 3 (*Task Type*: checking, detection, and response implementation) × 2 (*Experience Level*: novice and expert) ANOVAs were conducted for each of the four measures to determine if there was a significant difference between task types and between experience level. The interactive effect of task type and experience level on the four measures was also assessed. Task type was a repeated-measures variable and experience level was a between-participants variable.

For percent communications completed correctly, no main effect for task type was found, $(p = .15)$. However, a main effect was found for experience level for percent communications correctly completed by RO1 participants, $F(1, 74) = 7.69, p = .01, {\eta_p}^2 = .09$, (Figure 16) such that novice participants ($M = 84.33, SD = 24.79$) on average performed more 3-way communication events correctly compared to expert participants ($M = 58.80, SD = 33.94$). No interaction effect was found between task type and experience level for percent communications completed correctly, ($p = .09$).

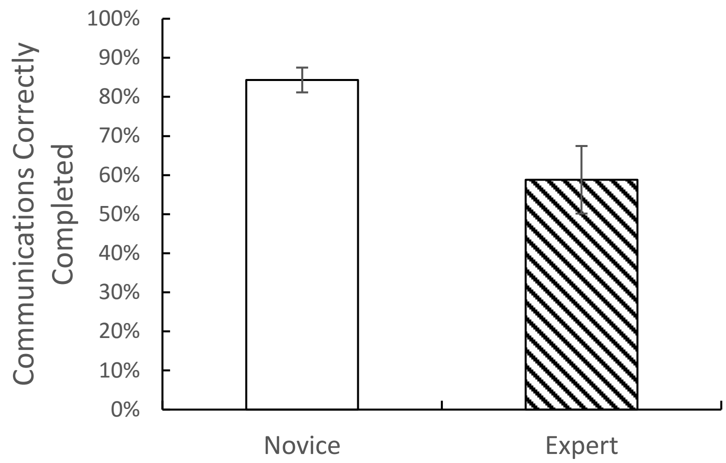

Figure 16: RO1 percent communications correctly completed by experience level. Error bars represent standard error.

For number of I&C location help requests, no main effect was found for task type, ($p = .71$), or experience level,($p = .49$). No interaction effect between the task type and experience level was found for the number of I&C location help requests, ($p = .71$).

For the number of clarifications required, a main effect was found for task type, $F(1.50, 110.79) = 3.51, p = .05, {\eta_p}^2 = .05$. However, the pairwise comparisons did not reveal any differences for the number of clarifications required for any of the task types. No main effect for experience level was found for the number of participant communications requiring clarification, ($p = .14$). Nor was there an interaction effect between the task type and experience level for the number of clarifications required, ($p = .31$).

For number of requests for repeating an instruction, a main effect was found for task type, $F(1.51,111.79) = 19.76, p < .01, {\eta_p}^2 = .21$, (Figure 17) such that detection ($M = 1.36, SD = 1.25$) had more repeat requests that both checking ($M = 0.28, SD = 0.48$) and response implementation ($M = 0.37, SD = 0.67$).

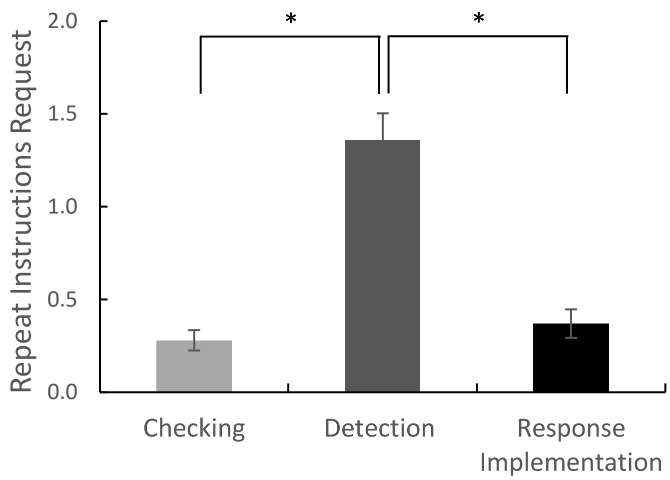

Figure 17: RO1 repeat instruction requests for the three task types. Error bars represent standard error. Asterisk represents statistical significance.

No main effect for experience level was found for the number of times a repeat instruction was requested, ($p = .08$). No interaction effect was found between the task type and experience level for the number of times a repeat instruction was requested, ($p = .87$).

Navigation and Identification

Navigation and identification variables included (i) number of correctly identified I&Cs, (ii) locating and identifying the correct I&C on the first attempt, (iii) the number of additional identifications made on the correct I&C, and (iv) the number of identifications made on I&Cs not relevant to the task. Four mixed 3 (*Task Type*: checking, detection, and response implementation)

× 2 (*Experience Level*: novice and expert) ANOVAs were conducted for each of the four measures to determine if there was a significant difference between task types and between experience level. The analyses also revealed if the novices and experts showed similar patterns of differences in performance across the task types. Task type was a repeated-measures variable and experience level was a between-participants variable.

For the number of correctly identified I&Cs, no main effect was found for task type, ($p =$ .53). A main effect for experience level was found number of correctly identified I&Cs, $F(1{,}76) = 4.99, p = .03, {\eta_p}^2 = .06$, (Figure 18) such that expert participants ($M = 3.81, SD = 0.24$) identified the correct I&C more often compared to the novice participants ($M = 3.17, SD = 0.86$). No interaction effect between the task type and experience level for the number of correctly identified I&Cs, ($p = .74$).

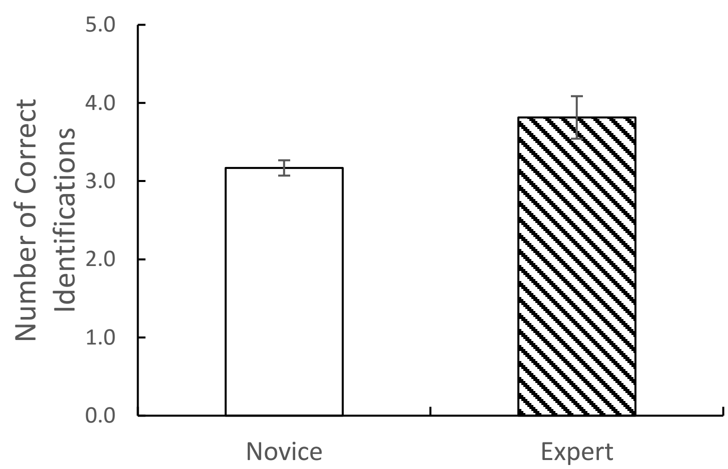

Figure 18: RO1 number of correct I&C identifications by experience level. Error bars represent standard error.

For locating and identifying the correct I&C on the first attempt, no main effect was found for task type, ($p = .26$), or experience level, ($p = .06$). No interaction effect was found between the task type and experience level for locating and identifying the correct I&C on the first attempt, ($p = .84$).

For the number of additional identifications made on the correct I&C, no main effect was found for task type, ($p = .09$), or experience level, ($p = .99$). No interaction effect was found between task type and experience level for the number of additional identifications made on the correct I&C, ($p = .97$).

For the number of identifications made on I&Cs not relevant to the task, no main effect was found for task type, ($p = .89$), or experience level, ($p = .61$). No interaction effect was found between task type and experience level for the number of identifications made on I&Cs not relevant to the task, ($p = .26$).

Action

Independent sample t-tests were conducted to determine if there were significant differences between novices and experts for various action performance variables. Below are the descriptions and results of each action performance measure for the detection and response implementation task types.

Detection

The percentage of correct gauge change detections, percentage of missed gauge change events, and the number of false positive detections for each participant were measured while completing the detection task. No differences were found for the percentage of correct gauge change detections for novice participants ($M = 48.52, SD = 25.26$) and expert participants ($M = 47.86, SD = 22.74$); ($p = .94$). No differences were found in percentage of missed gauge change events between novices ($M = 42.31, SD = 24.49$) and experts ($M = 43.50, SD = 23.23$); ($p = .89$). No differences were found in the number of false positive detections for novice participants ($M = 122.45, SD = 567.05$) and expert participants ($M = 61.11, SD = 45.71$); ($p = .75$).

Response Implementation

The number of times a participant followed the correct sequence of identifying the I&C and then manipulating it in the correct direction, the percentage of correct manipulations, percentage of description error, and the percentage of mode errors for each participant was measured while completing the response implementation task. No differences in the number of correctly followed sequences was found for novice participants ($M = 0.35, SD = 0.48$) and expert participants ($M = 0.44, SD = 0.53$); ($p = .58$). A significant difference in correct manipulations was found between novice participants ($M = 65.37, SD = 33.10$) and expert participants ($M = 83.57, SD = 17.44$), such that experts performed the correct (open/close) manipulation more often compared to novices; $t(16.85) = -2.58, p = .02$ (Figure 19).

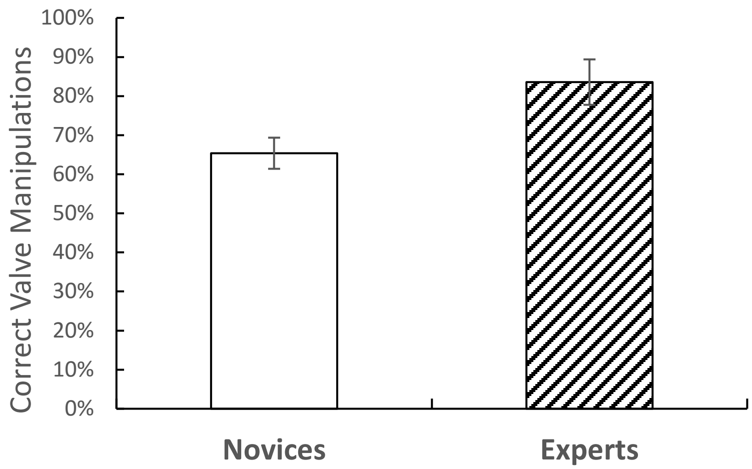

Figure 19: RO1 percentages of correctly manipulated I&Cs during the response implementation task type by experience level. Error bars represent standard error.

No differences in the number of description error manipulations was found for novice participants ($M = 0.21, SD = 1.72$) and expert participants ($M = 4.37, SD = 9.07$); ($p = .21$). A significant difference was found for mode error manipulations for novice participants ($M = 19.79, SD = 28.53$) and expert participants ($M = 2.22, SD = 6.67$), such that experts had a lower mode error percentage than novice participants; $t(54.97) = 4.29, p < .01$ (Figure 20).

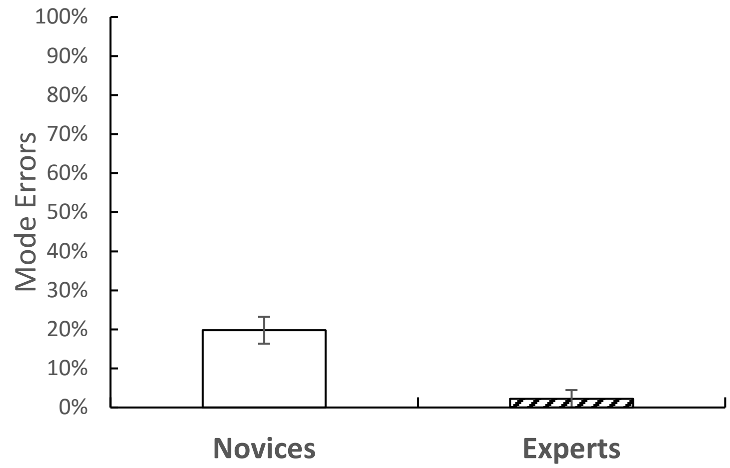

Figure 20: RO1 percentages of manipulated that were mode error during the response implementation task type by experience level. Error bars represent standard error.

Physiological Measures

All dependent variables that were used in the ANOVAs were calculated by taking a difference from a five-minute wakeful resting baseline. For example, if the participant's heart rate for the five-minute baseline was 53 beats per minute (BPM) and their heart rate for the subsequent detection task was 62 BPM, their difference from baseline would be 9 BPM, $(62 - 53 = 9)$. This approach helps account for individual differences when comparing group means in an ANOVA.

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) as well as lobes (i.e., compare brain activity among the frontal, parietal and occipital lobes). Hence, a series of ANOVAs were performed.

Hemispheres

Six mixed 3 (*Task Type*: checking, detection, and response implementation) × 2 (*Experience Level*: experienced and expert) ANOVAs were run for left and right hemispheres for theta, alpha, and beta frequency bands. These ANOVAs provided insight in to the overall effects of task type and experience level on the left and right hemispheres.

Theta in the Left and Right Hemispheres

For theta in the left hemisphere, a main effect was found for task type, $F(1.35, 100.94) = 4.14, p = .03, {\eta_p}^2 = .05$, (Figure 21) such that detection ($M = -1.87, SD = 859.03$) had less of a change from baseline than response implementation ($M = 534.12, SD = 1122.78$), and checking ($M = 537.48, SD = 1518.21$) was not different than either detection or response implementation. No main effect was found for experience level for theta in the left hemisphere, ($p = .08$). No interaction effect between task type and experience level for theta left hemisphere, ($p = .72$).

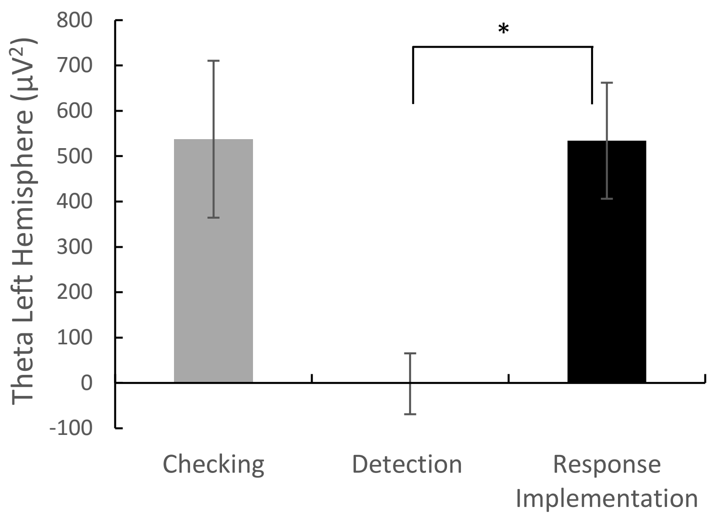

Figure 21: RO1 theta left hemisphere difference from baseline in μV^2 for the three task types. Error bars represent standard error. Asterisk represents statistical significance.

For theta recorded in the right hemisphere, a main effect was found for task type, $F(2,150) = 3.83, p = .02, {\eta_p}^2 = .05$, (Figure 22) such that detection ($M = 74.40, SD = 548.70$) had less of a change from baseline than response implementation ($M = 492.73, SD = 867.75$), and checking ($M = 414.12, SD = 953.11$) did not differ from either the detection or response implementation. No main effect for experience level was found for theta in the right hemisphere, ($p = .21$). No interaction effect was found between the task type and experience level for theta in the right hemisphere, ($p = .49$).

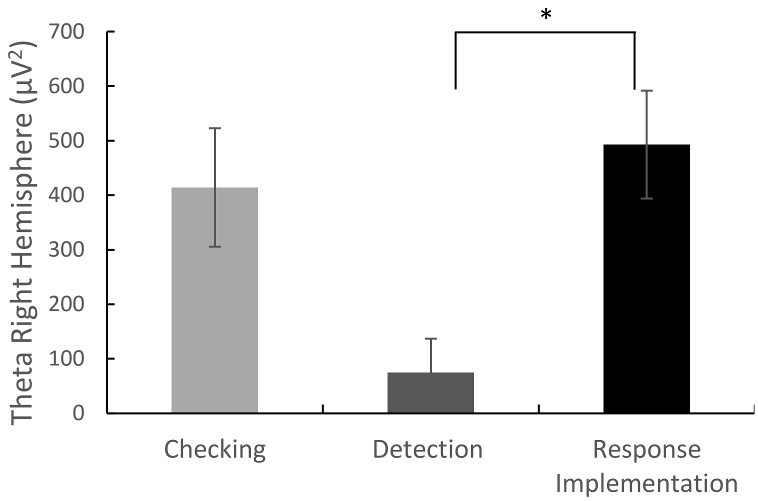

Figure 22: RO1 theta right hemisphere difference from baseline in μV^2 for the three task types. Error bars represent standard error. Asterisk represents statistical significance.

Alpha in the Left and Right Hemispheres

For alpha in the left hemisphere, no main effect was found for task type, ($p = .21$), or experience level, ($p = .22$). No interaction effect between the task type and experience level was found for alpha in the left hemisphere, ($p = .66$).

For alpha in the right hemisphere, no main effect was found for task type, ($p = .25$), or experience level, ($p = .10$). No interaction effect between the task type and experience level was found for alpha in the right hemisphere, ($p = .21$).

Beta in the Left and Right Hemispheres

For beta in the left hemisphere, a main effect was found for task type, $F(2{,}150) = 5.50, p < .01, {\eta_p}^2 = .07$, (Figure 23) such that detection ($M = 587.27, SD = 1572.99$) had less of a change from baseline than response implementation ($M = 2201.80, SD = 2126.57$), and checking ($M = 1869.61, SD = 1784.17$) did not differ from either detection or response implementation.

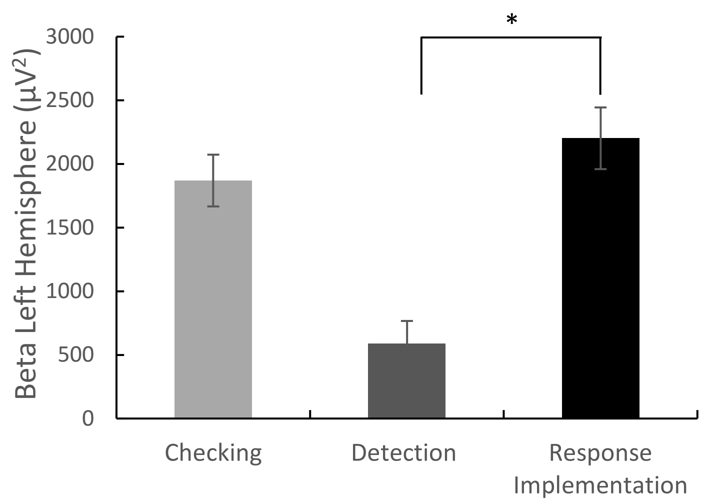

Figure 23: RO1 beta left hemisphere difference from baseline in μV^2 for the three task types. Error bars represent standard error. Asterisk represents statistical significance.

No main effect for experience level was found for beta in the left hemisphere, ($p = .22$). An interaction effect was found between task type and experience level for beta in the left hemisphere, $F(2{,}150) = 3.58, p = .03, {\eta_p}^2 = 05$. However, simple effects did not reveal any significant differences in beta in the left hemisphere between novices and experts for any of the task types.

For beta in the right hemisphere, a main effect was found for task type, $F(1.54{,}115.67) = 4.50, p = .02, {\eta_p}^2 = .06$, (Figure 24) such that detection ($M = 594.13, SD = 2044.50$) had less of a change from baseline than both checking ($M = 1780.04, SD = 2259.62$) and response implementation ($M = 2247.76, SD = 3078.62$). No main effect was found for experience level for beta in the right hemisphere, $F(1{,}75) =$

$1.61, p = .21, \eta_p{}^2 = .02$. No interaction effect was found between the task type and experience level for beta in the right hemisphere, $(p = .12)$.

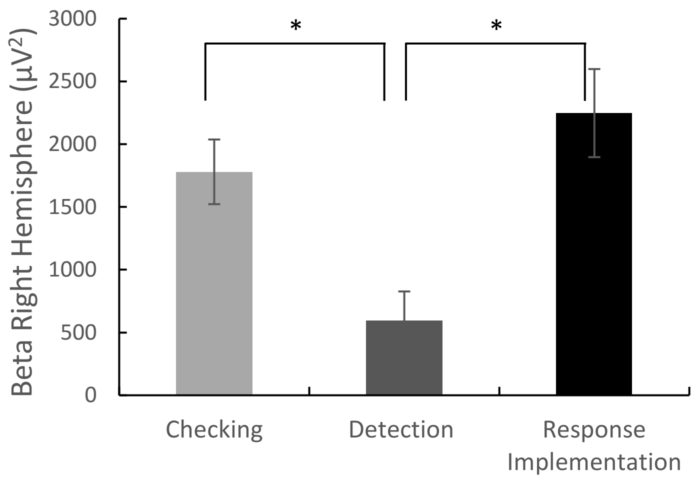

Figure 24: RO1 beta right hemisphere difference from baseline in μV^2 for the three task types. Error bars represent standard error. Asterisk represents statistical significance.

Lobes

Nine mixed 3 (*Task Type*: checking, detection, and response implementation) × 2 (*Experience Level*: novice and expert) ANOVAs were run for frontal, parietal and occipital lobes for theta, alpha, and beta frequency bands. These ANOVAs provided insight in to the overall effects of task type and experience level on the frontal, parietal and occipital lobes.

Theta in the Frontal, Parietal, and Occipital Lobes

For theta in the frontal lobe, no main effect was found for task type, $(p = .26)$, or experience level, $(p = .06)$. Also, no interaction effect was found between the task type and experience level for theta in the frontal lobe, $(p = .28)$.

For theta in the parietal lobe, no main effect was found for task type, ($p = .21$), or experience level, ($p = .18$). Also, no interaction effect was found between task type and experience level for theta in the parietal lobe, ($p = .58$).

For theta in the occipital lobe, no main effect was found for task type, ($p = .31$), or experience level, ($p = .16$). Also, no interaction effect between the task type and experience level was found for theta in the occipital lobe, ($p = .63$).

Alpha in the Frontal, Parietal, and Occipital Lobes

For alpha in the frontal lobe, no main effect was found for task type, ($p = .09$). or experience level, ($p = .09$). No interaction effect between the task type and experience level was found for alpha in the frontal lobe, ($p = .09$).

For alpha in the parietal lobe, no main effect was found for task type, ($p = .13$), or experience level, ($p = .12$). No interaction effect between the task type and experience level was found for alpha in the parietal lobe, ($p = .06$).

For alpha in the occipital lobe, no main effect was found for task type, ($p = .28$), or experience level, ($p = .19$). No interaction effect between task type and experience level was found for alpha in the occipital lobe, ($p = .86$).

Beta in the Frontal, Parietal, and Occipital Lobes

For beta in the frontal lobe, a main effect was found for task type, $F(2{,}150) = 4.53, p = .01, {\eta_p}^2 = .06$, (Figure 25) such that detection ($M = 5.92.93, SD = 576.74$) had less of a change from baseline than both checking ($M = 1842.44, SD = 1789.43$) and response implementation ($M = 2142.07, SD = 2245.49$).

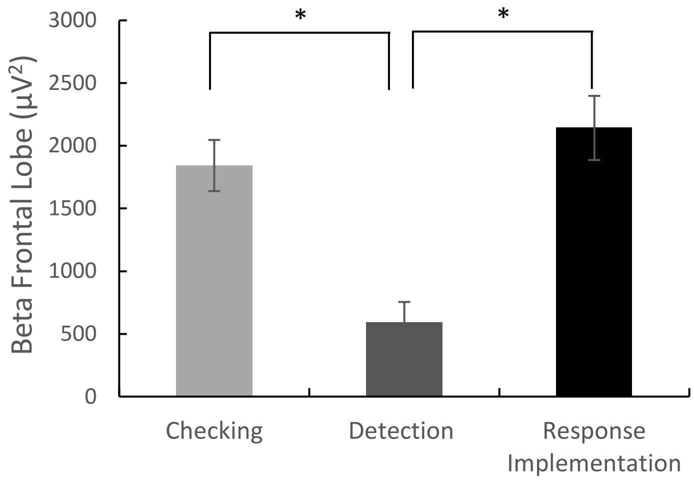

Figure 25: RO1 beta frontal lobe difference from baseline in μV^2 for the three task types. Error bars represent standard error. Asterisk represents statistical significance.

No main effect was found for experience level, $(p = .30)$. An interaction effect was found between task type and experience level for beta in the frontal lobe, $F(1{,}150) = 4.14, p = .02, {\eta_p}^2 = .05$. However, the simple effects did not reveal any significant differences in beta in the frontal lobe between novices and experts for any of the task types.

For beta in the parietal lobe, a main effect was found for task type, $F(2{,}150) = 5.61, p < .01, {\eta_p}^2 = .07$, (Figure 26) such that detection $(M = 404.03, SD = 1549.04)$ had less of a change from baseline that both checking $(M = 1612.40, SD = 1834.84)$ and response implementation $(M = 1891.00, SD = 2001.09)$.

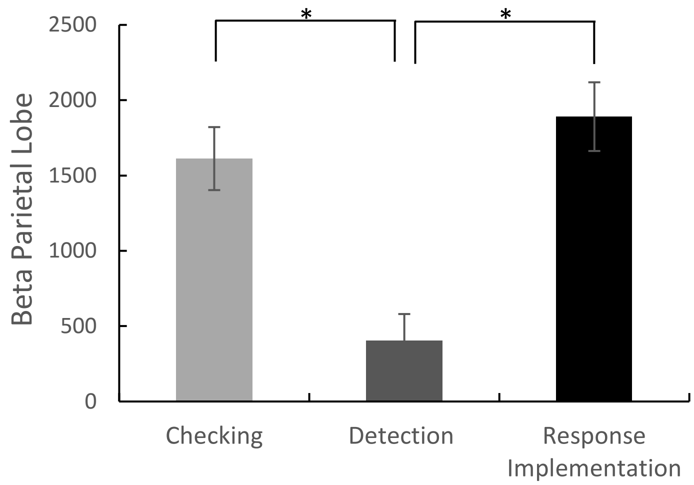

Figure 26: RO1 beta parietal lobe difference from baseline in μV^2 for the three task types. Error bars represent standard error. Asterisk represents statistical significance.

No main effect for experience level was found for beta in the parietal lobe, ($p = .12$). An interaction effect was found between the task type and experience level for beta in the parietal lobe, $F(1{,}150) = 3.48, p = .03, {\eta_p}^2 = .04$. However, the simple effects did not reveal any significant differences in beta in the parietal lobe between novices and experts for any of the task types.

For beta in the occipital lobe, a main effect was found for task type, $F(1.57{,}117.89) = 6.40, p < .01, {\eta_p}^2 = .08$, (Figure 27) such that detection ($M = 524.00, SD = 1687.14$) had less of a change from baseline than response implementation ($M = 1634.54, SD = 2361.55$), and checking ($M = 1152.98, SD = 1920.76$) did not differ from either detection or response implementation. No main effect for experience level was found for beta in the occipital lobe, ($p = .18$). Also, no interaction effect was found between task type and experience level for beta in the occipital lobe, ($p = .44$).

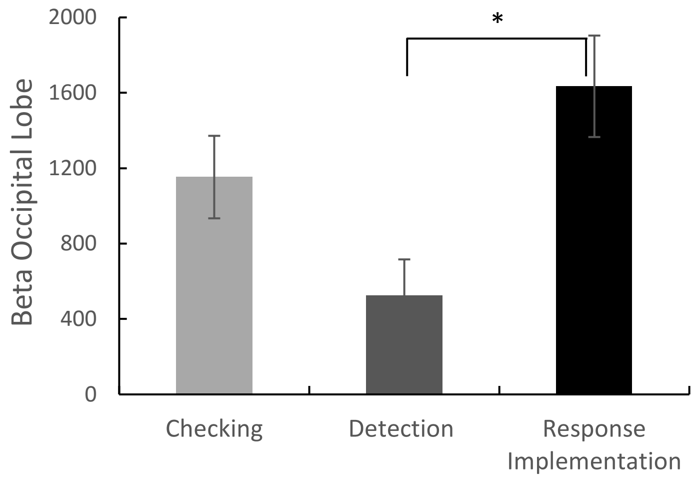

Figure 27: RO1 beta occipital lobe difference from baseline in μV^2 for the three task types. Error bars represent standard error. Asterisk represents statistical significance.

Electrocardiogram (ECG)

Three mixed 3 (*Task Type*: checking, detection, and response implementation) × 2 (*Experience Level*: novice and expert) ANOVAs were conducted to determine if the different task types and experience levels impacted heart rate (HR), interbeat interval (IBI), and/or heart rate variability (HRV). These analyses also assessed the interactive effects between the task types and experience level, which would reveal if any of the observed differences across task types were similar for novices and experts. Task type was a repeated-measures variable and experience level was a between-participants variable. All measures were calculated as a difference from baseline to mitigate individual differences.

Heart rate (HR)

HR was derived from R-Peak detections using the So-Chan QRS algorithm from a raw electrical ECG signal and is measured in BPM. A main effect for task type was found for HR,

$F(1.84{,}137.84) = 19.34, p < .01, {\eta_p}^2 = .21$, (Figure 28) such that detection ($M = -5.5, SD = 20.93$) had a larger decrease from baseline that both checking ($M - 1.33, SD = 13.51$), and response implementation ($M = -4.48, SD = 15.86$) for HR.

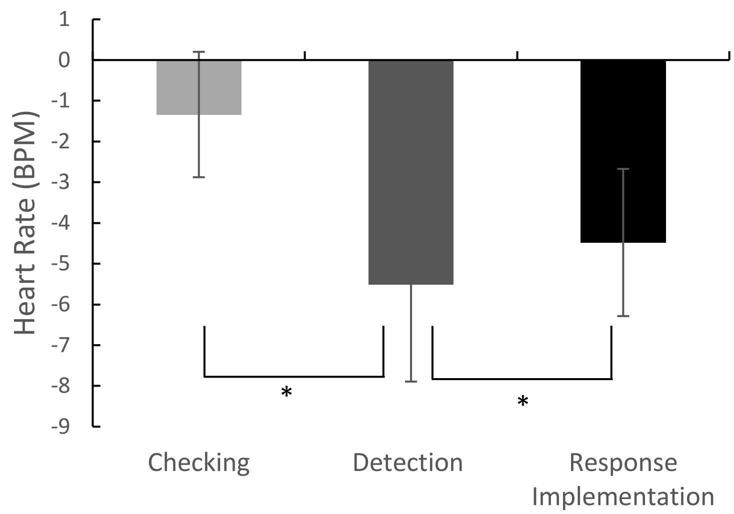

Figure 28: RO1 HR difference from baseline in BPM for the three task types. Error bars represent standard error. Asterisk represents statistical significance.

No main effect was found for experience level ,($p = .08$). An interaction effect between the task type and experience level was found for HR, $F(2{,}150) = 6.24, p < .01, {\eta_p}^2 = .08$. However, simple effects did not reveal any differences in HR between novices and experts for any of the task types.

Interbeat Interval (IBI)

IBI was derived from R-Peak detections using the So-Chan QRS algorithm from a raw ECG electrical signal. For IBI, a main effect for task type was found, $F(2{,}150) = 4.16, p = .02, {\eta_p}^2 = .05$, (Figure 29) such that checking ($M = -21.21, SD = 54.50$) was lower than response implementation ($M = -15.14, SD = 57.40$), and detection ($M = 414.12, SD = 953.11$) did not differ from either checking or response implementation type.

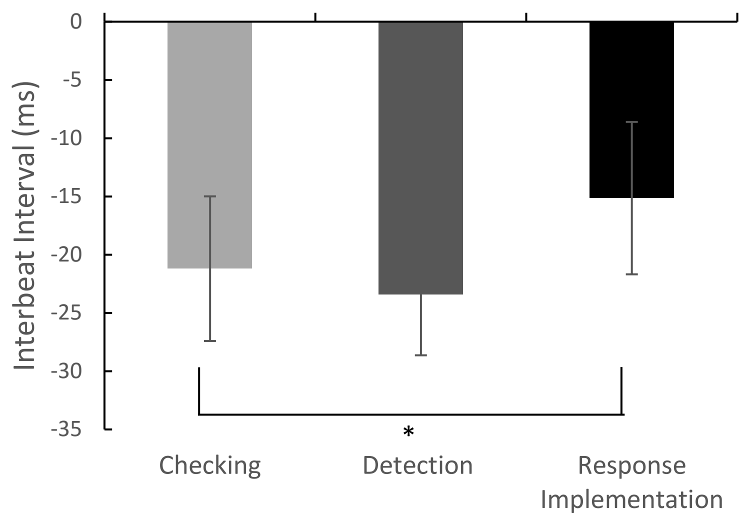

Figure 29: RO1 IBI difference from baseline in milliseconds for the three task types. Error bars represent standard error. Asterisk represents statistical significance.

No main effect was found for experience level for IBI, ($p = .80$). No interaction effect was found between task type and experience level for IBI, ($p = .07$).

Heart Rate Variability (HRV)

For HRV, a main effect for task type was found, $F(2,150) = 3.75, p = .03, {\eta_p}^2 = .05$, (Figure 30) such that detection ($M = 4.99, SD = 20.787$) had less of a change from baseline than response implementation ($M = 20.15, SD = 22.15$), and checking ($M = 13.43, SD = 21.14$) did not differ from either detection or response implementation type.

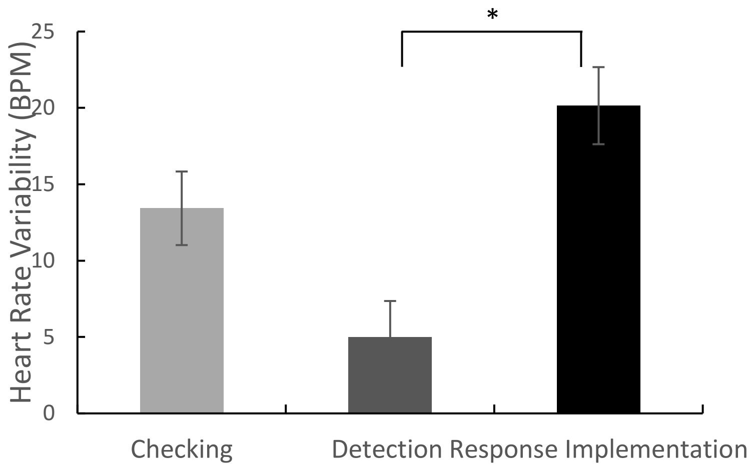

Figure 30: RO1 HRV difference from baseline in BPM for the three task types. Error bars represent standard error. Asterisk represents statistical significance.

No main effect for experience level was found for HRV, ($p = .57$). No interaction effect was found between task type and experience level for HRV, ($p = .08$).

RO2 Analysis (Experienced and Expert)

Subjective Measures

Instantaneous Self-Assessment

A mixed 3 (*Task Type*: checking, detection, and response implementation) × 2 (*Experience Level*: experienced and expert) ANOVA was conducted. Task type was a repeated measures factor while experience level was a between-participants factor. The ANOVA was used to determine if task type and experience level have significant effects on online-subjective workload (i.e., reported as the task was being performed), and to determine if the pattern of workload differences found across the task types differed between experienced participants and expert participants. No main effect found for task type was found for ISA, ($p = .27$). A main effect for experience level was found for ISA, $F(1,19) = 20.64, p < .01, {\eta_p}^2 = .52$, (Figure 31) such that experienced participants ($M = 1, SD = 0$) reported lower ISA compared to expert participants ($M = 1.63, SD = 0.48$). No interaction effect between the task type and experience level was found for ISA, ($p = .27$).

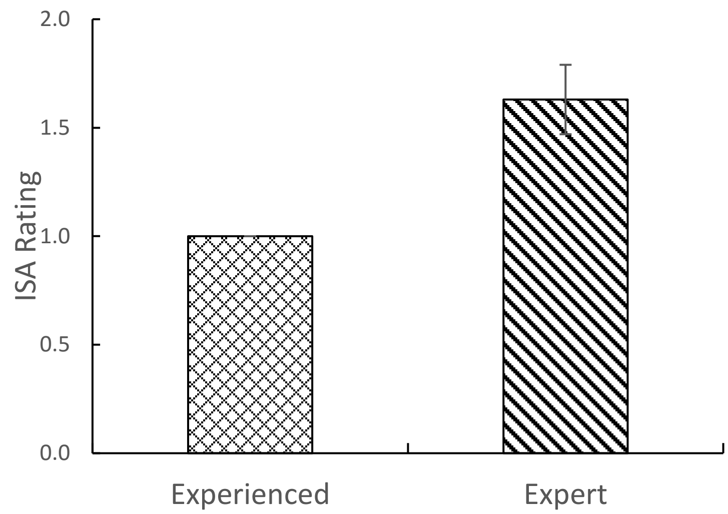

Figure 31: RO2 mean ISA by experience level. Error bars represent standard error.

NASA Task Load Index (TLX)

A mixed 3 (*Task Type*: checking, detection, and response implementation) × 2 (*Experience Level*: experienced and expert) ANOVA was conducted for TLX global workload as well as each of the six subscales of the NASA TLX (mental demand, physical demand, temporal demand, performance, effort, and frustration). Task type was a repeated-measure factor, and experience level was a between-participants variable. The ANOVAs were used to determine if there was a significant workload difference between task types, and to see if there were differences in the ratings between experienced and expert participants.

Task type had a significant effect on the TLX global workload and several of the subscales that persisted across experience level. A main effect for task type was found for TLX global workload, $F(2,38) = 5.01, p = .01, {\eta_p}^2 = .21$, such that detection ($M = 19.96, SD = 26.92$) was higher than both checking ($M = 15.55, SD = 20.74$) and response implementation ($M = 13.69, SD = 19.86$). A main effect for task type was also found for physical demand, $F(2,38) = 8.77, p < .01, {\eta_p}^2 = .36$, such that detection ($M = 15.00, SD = 24.95$) elicited higher physical demand than the checking ($M = 4.05, SD = 8.46$), and response implementation ($M = 8.81, SD = 19.03$) did not differ from either the checking or detection. A main effect for task type was found for performance, $F(2,38) = 3.95, p = .03, {\eta_p}^2 = .17$, such that detection ($M = 24.29, SD = 36.20$) elicited higher performance demands than response implementation ($M = 12.86, SD = 24.52$), and checking ($M = 17.86, SD = 26.86$) did not differ from either the detection or response implementation.

Experience level showed significant effects on the TLX global workload as well as all subscales that persisted across task type. A main effect for experience level was found for TLX global workload, $F(1,19) = 66.18, p < .01, {\eta_p}^2 = .77$, such that experienced participants ($M = 0.26, SD = 0.49$) reported lower global workload compared to expert participants ($M = 37.93, SD = 16.18$). A main effect for experience level was found for mental demand, $F(1, 19) = 54.65, p < .01, {\eta_p}^2 = .74$, (Figure 32) such that experienced participants ($M = 0, SD = 0$) reported lower mental demand compared to expert participants ($M = 49.07, SD = 23.20$).

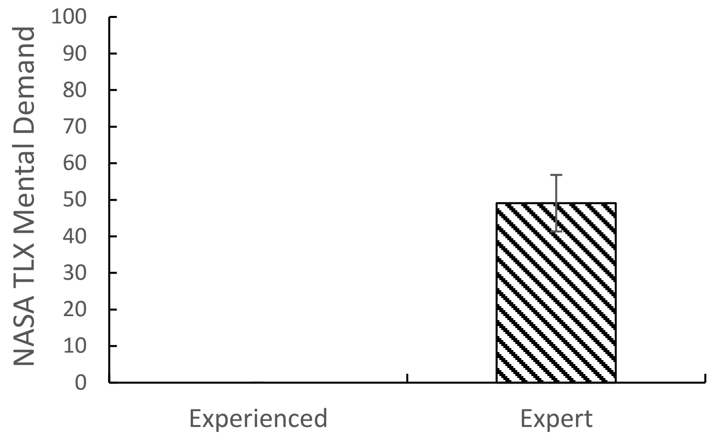

Figure 32: RO2 NASA TLX mean mental demand by experience level. Error bars represent standard error.

A main effect for experience level was found for the physical demand, $F(1, 76) = 4.06, p = .047, {\eta_p}^2 = .05$, such that experienced participants ($M = 0, SD = 0$) reported lower physical demand compared to expert participants ($M = 21.67, SE = 18.86$). A main effect for experience level was found for the temporal demand, $F(1, 19) = 36.30, p < .01, {\eta_p}^2 = .66$, (Figure 33) such that experienced participants ($M = 0.56, SD = 1.30$) reported lower temporal demand compared to expert participants ($M = 29.44, SD = 16.69$).

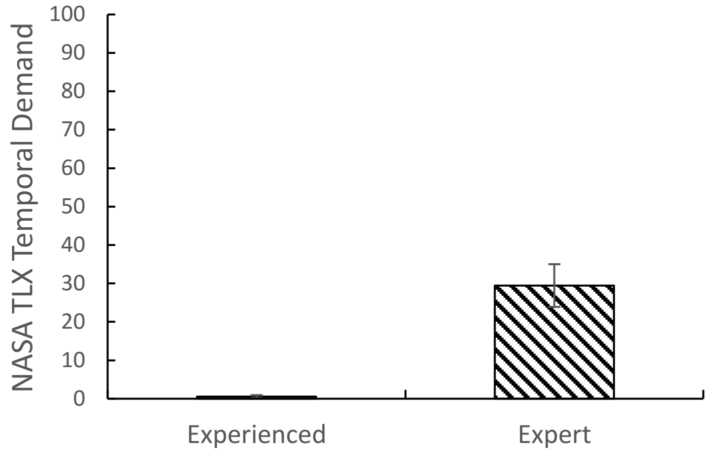

Figure 33: RO2 NASA TLX mean temporal demand by experience level. Error bars represent standard error.

A main effect for experience level was found for performance, $F(1, 19) = 39.01, p < .01, {\eta_p}^2 = .67$, such that experienced participants ($M = 0, SD = 0$) reported lower performance compared to expert participants ($M = 42.78, SE = 23.93$). A main effect for experience level was found for effort, $F(1, 19) = 37.21, p < .01, {\eta_p}^2 = .66$, such that experienced participants ($M = 0.83, SD = 1.51$) reported lower effort compared to expert participants ($M = 41.48, SD = 23.22$). A main effect for experience level was found for frustration, $F(1, 19) = 46.96, p < .01, {\eta_p}^2 = .71$, (Figure 34) such that experienced participants ($M = 0.14, SD = 0.48$) reported lower frustration compared to expert participants ($M = 43.15, SD = 21.93$).

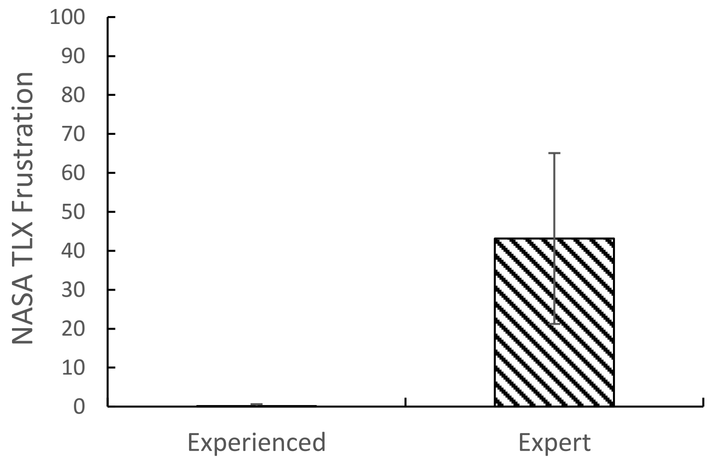

Figure 34: RO2 NASA TLX mean frustration by experience level. Error bars represent standard error.

Interaction effects between task type and experience level were found for TLX global workload and several of the subscales. An interaction effect was found between task type and experience level for TLX global workload, $F(2, 38) = 5.11, p = .01, {\eta_p}^2 = .21$, (Figure 35). Checking, $F(1,8) = 49.72, p < .01, d = 3.62$, elicited lower global workload for experienced participants ($M = 0.07, SD = 0.24$) than for expert participants ($M = 36.20, SD = 15.37$). Detection, $F(1, 8.02) = 41.68, p < .01, d = 3.32$, elicited lower global workload for experienced participants ($M = 0.28, SD = 0.96$) than for expert participants ($M = 46.20, SD = 21.32$). Response implementation, $F(1,8.04) = 23.42, p < .01, d = 2.48$, elicited lower global workload for experienced participants ($M = 0.42, SD = 1.04$) than for expert participants ($M = 13.69, SD = 19.86$).

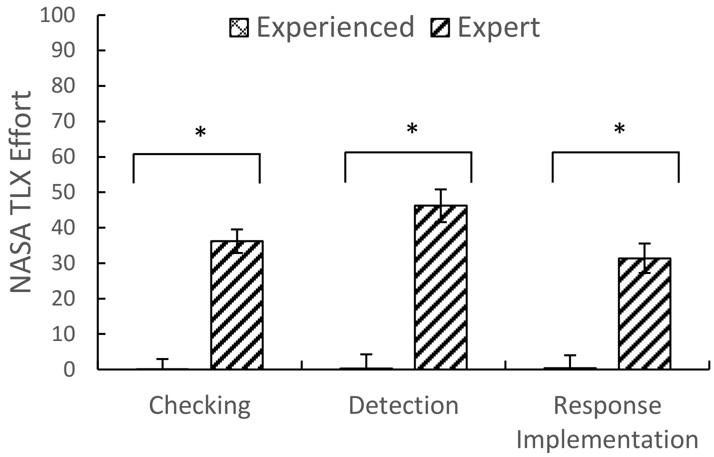

Figure 35: RO2 NASA TLX mean global workload by task type and experience level. Error bars represent standard error. Asterisk represents statistical significance.

A significant interaction effect was found between task type and experience level for physical demand, $F(2,38) = 8.77, p < .01, {\eta_p}^2 = .32$, (Figure 36). Checking, $F(1,20) = 8.97, p < .01, d = 1.32$, elicited lower physical demand for experienced participants ($M = 0, SD = 0$) than for expert participants ($M = 9.44, SD = 11.02$). Detection, $F(1,20) = 19.46, p < .01, d = 1.95$, elicited lower physical demand for experienced participants ($M = 0, SD = 0$) than for expert participants ($M = 35, SD = 27.73$). Response implementation, $F(1,20) = 8.14, p = .01, d = 1.26$, elicited lower physical demand for experienced participants ($M = 0, SD = 0$) than for expert participants ($M = 20.56, SD = 25.18$).

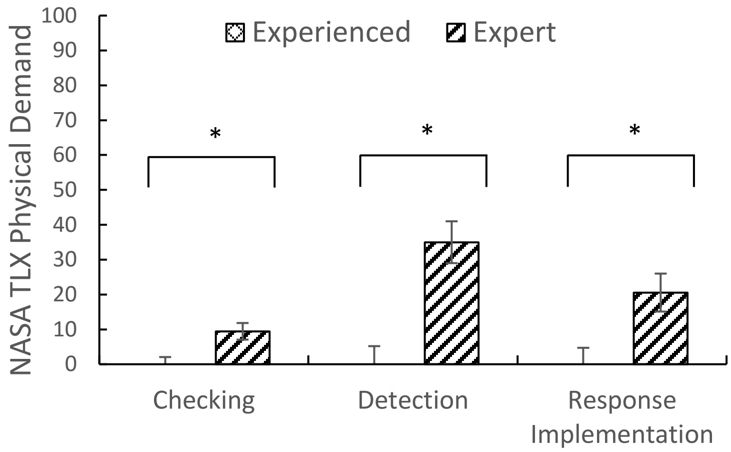

Figure 36: RO2 NASA TLX mean physical demand by task type and experience level. Error bars represent standard error. Asterisk represents statistical significance.

An interaction effect was found between task type and experience level for performance, $F(2{,}38) = 3.95, p < .03, {\eta_p}^2 = .17$, (Figure 37). Checking, $F(1{,}20) = 30.84, p < .01, d = 2.45$, elicited lower performance for experienced participants ($M = 0, SD = 0$) than for expert participants ($M = 41.67, SD = 26.22$). Detection, $F(1{,}20) = 32.35, p < .01, d = 2.51$, elicited lower performance for experienced participants ($M = 0, SD = 0$) than for expert participants ($M = 56.67, SD = 34.82$). Response implementation, $F(1{,}20) = 11.88, p < .01, d = 1.52$, elicited lower performance for experienced participants ($M = 0, SD = 0$) than for expert participants ($M = 30.00, SD = 30.41$).

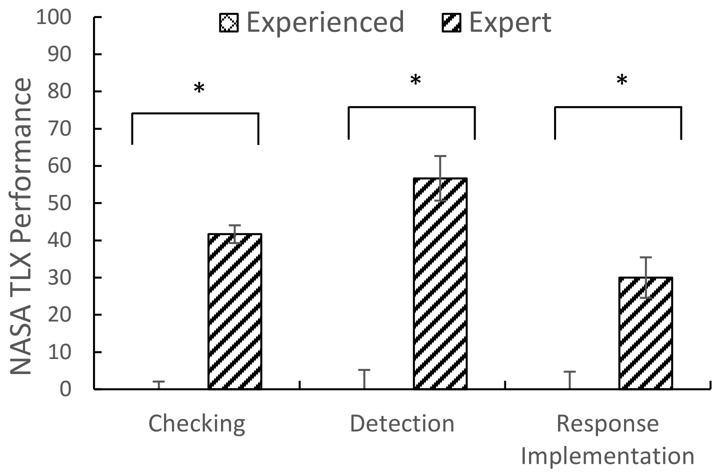

Figure 37: RO2 NASA TLX mean performance by task type and experience level. Error bars represent standard error. Asterisk represents statistical significance.

An interaction effect was found between task type and experience level for effort, $F(2{,}38) = 3.81, p = .03, {\eta_p}^2 = .17$, (Figure 38). Checking, $F(1{,}20) = 46.28, p < .01, d = 3.00$, elicited lower effort for experienced participants ($M = 0, SD = 0$) than for expert participants ($M = 46.11, SD = 23.69$). Detection, $F(1{,}8.11) = 20.39, p < .01, d = 2.31$, elicited lower effort for experienced participants ($M = 0.83, SD = 2.89$) than for expert participants ($M = 46.11, SD = 29.98$). Response implementation, $F(1{,}8.30) = 13.19, p < .01, d = 1.85$, elicited lower effort for experienced participants ($M = 1.67, SD = 3.89$) than for expert participants ($M = 32.22, SD = 25.01$).

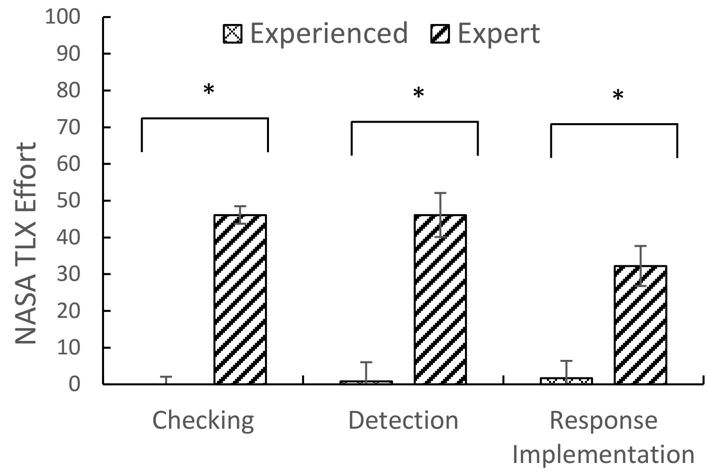

Figure 38: RO2 NASA TLX mean effort by task type and experience level. Error bars represent standard error. Asterisk represents statistical significance.

Performance Measures

Communication Reporting

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. Four mixed 3 (*Task Type*: checking, detection, and response implementation) × 2 (*Experience Level*: experienced and expert) ANOVAs were conducted for each of the communication reporting variables to determine if there was a significant difference between task type and between experience level. The interactive effect of task type and experience level on the four measures was also assessed. Task type was a repeated-measures variable and experience level was a between-participants variable.

For percent communications completed correctly, no main effect was found for task type, $(p = .08)$. A main effect was found for experience level for percent communications completed correctly, $F(1,19) = 8.28, p = .01, {\eta_p}^2 = .30$, (Figure 39) such that experienced participants

$(M = 100, SD = 0)$ on average performed more 3-way communication events correctly compared to expert participants $(M = 68.06, SD = 38.81)$.

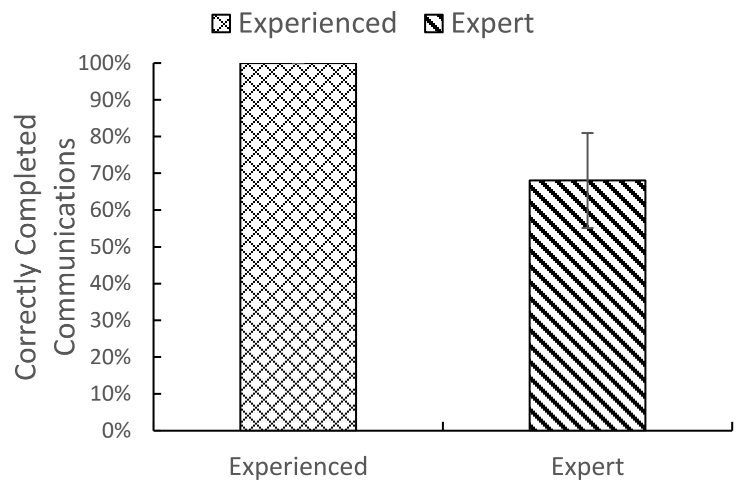

Figure 39: RO2 percentage of correctly completed communications by task type and experience level. Error bars represent standard error.

No interaction effect was found between task type and experience level for percent communications completed correctly, $(p = .08)$.

For number of I&C location help requests, no main effect was found for task type, $(p = .13)$, or experience level, $(p = .12)$. No interaction effect was found between task type and experience level for number of I&C location help requests, $(p = .13)$.

For the number of clarifications required, no main effect was found for task type, $(p = .32)$. However, a main effect was found for experience level for the number of clarifications required, $F(1,19) = 11.21, p < .01, {\eta_p}^2 = .37$, (Figure 40) such that experienced participants $(M = 0, SD = 0)$ required fewer clarification compared to expert participants $(M = 0.30, SD = 0.31)$.

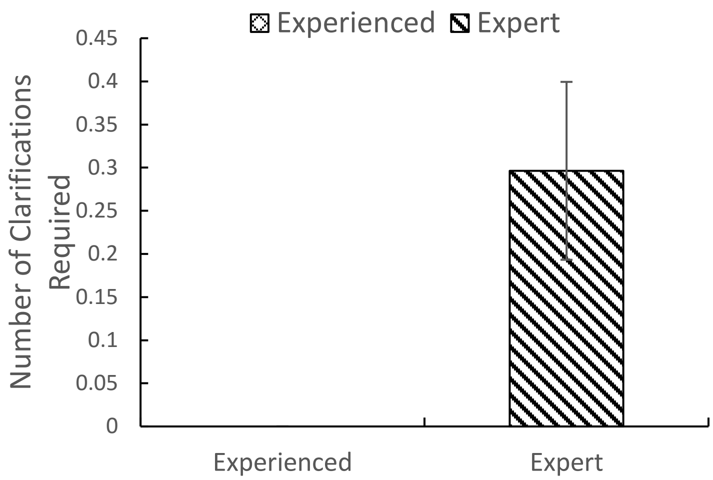

Figure 40: RO2 number of clarifications required. Error bars represent standard error.

No interaction effect was found between the task type and experience level for the number of clarifications required, ($p = .32$).

For number of requests for repeating an instruction, a main effect was found for task type, $F(1.39,26.46) = 6.49, p = .01, {\eta_p}^2 = .26$, such that detection ($M = 0.62\ SD = 1.02$) had more repeat requests than checking ($M = 0.24, SD = 0.70$), and response implementation ($M = 0.24, SD = 0.89$) did not differ from either checking or detection. A main effect was found for experience level for the number of requests for repeating an instruction, $F(1,19) = 8.60, p < .01, {\eta_p}^2 = .31$, such that experienced participants ($M = 0, SD = 0$) requested fewer repeat instructions compared to expert participants ($M = 0.85, SD = 1.02$). An interaction effect was found between task type and experience level for the number of requests for repeating an instruction, $F(1.39,26.46) = 6.49, p = .01, {\eta_p}^2 = .26$, (Figure 41). For checking, ($p = .07$), not observed difference was observed for number of repeat requests between experienced participants ($M = 0, SD = 0$) and expert participants ($M = 0.56, SD =$

1.01). Detection, $F(1,20) = 19.94, p < .01, d = 1.96$, had fewer occurrences of repeat requests for experienced participants ($M = 0, SD = 0$) than for expert participants ($M = 1.44, SD = 1.13$). For response implementation, ($p = .16$), not observed difference was observed for number of repeat requests between experienced participants ($M = 0, SD = 0$) than for expert participants ($M = 0.56, SD = 1.33$).

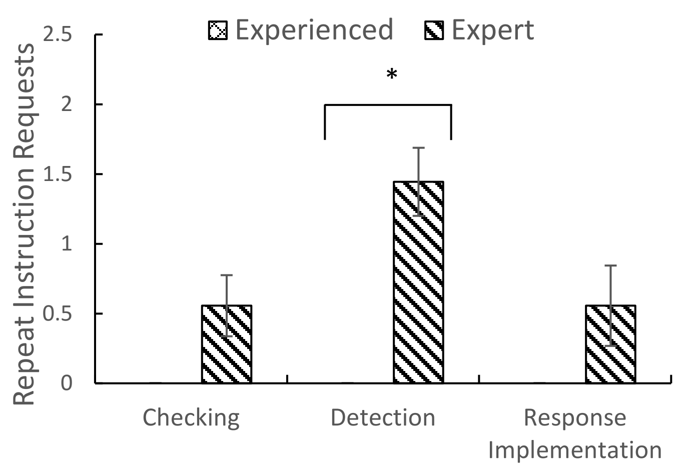

Figure 41: RO2 repeat instruction requests by task type and experience level. Error bars represent standard error. Asterisk represents statistical significance.

Navigation and Identification

Navigation and identification variables included (i) number of correctly identified I&Cs, (ii) locating and identifying the correct I&C on the first attempt, (iii) the number of additional identifications made on the correct I&C, and (iv) the number of identifications made on I&Cs not relevant to the task. Four mixed 3 (*Task Type*: checking, detection, and response implementation) × 2 (*Experience Level*: experienced and expert) ANOVAs were conducted for each of navigation and identification variables to determine if there was a significant difference between task type and experience level. The analyses also revealed if the experienced and expert participants showed similar patterns of differences in performance across the task type. Task type was a repeated-measures variable and experience level was a between-participants variable.

For the number of correctly identified I&Cs, a significant main effect was found for task type, $F(2,38) = 39.89, p < .01, {\eta_p}^2 = .68$, such that detection ($M = 2.38, SD = 1.75$) was lower than both checking ($M = 4, SD = 0$) and response implementation ($M = 4, SD = 0$). A main effect was found for experience level for the number of correctly identified I&Cs, $F(1,19) = 39.88, p < .01, {\eta_p}^2 = .68$, such that experienced participants ($M = 3.06, SD = 0.45$) had fewer correctly identified I&Cs compared to expert participants ($M = 4, SD = 0$). An interaction effect was found between task type and experience level for the number of correctly identified I&Cs, $F(2,38) = 39.89, p < .01, {\eta_p}^2 = .68$, (Figure 42). Checking had 100% correctly identified I&Cs by both the experienced participants ($M = 4, SD = 0$) and expert participants ($M = 4, SD = 0$). Detection $F(1,20) = 39.89, p < .01, d = 3.00$, had fewer correct I&Cs identified for experienced participants ($M = 1.17, SD = 1.33$) than for expert participants ($M = 4, SD = 0$). Response implementation task type had 100% correctly identified I&Cs by both the experienced participants ($M = 4, SD = 0$) and expert participants ($M = 4, SD = 0$).

Figure 42: RO2 number of correct identifications by task type and experience level. Each task type has four identifications possible. Error bars represent standard error. Asterisk represents statistical significance.

For locating and identifying the correct I&C on the first attempt, a main effect was found for task type, $F(1.12, 21.22) = 36.37, p < .01, {\eta_p}^2 = .66$, such that detection ($M = 2.14, SD = 1.74$) had a fewer occurrence of correct first attempts than both checking ($M = 3.90, SD = 0.30$) and response implementation ($M = 3.90, SD = 0.30$). A main effect was found for experience level for locating and identifying the correct I&C on the first attempt, $F(1,19) = 22.29, p < .01, {\eta_p}^2 = .54$, such that experienced participants ($M = 2.94, SD = 0.49$) had a fewer occurrence of correct first identification compared to expert participants($M = 3.81, SD = 0.29$). A significant interaction effect was found between task type and experience level for locating and identifying the correct I&C on the first attempt, $F(1.12, 21.22) = 26.80, p < .01, {\eta_p}^2 = .59$, (Figure 43). Checking, $F(1,20) = 0.04, p = .84, d = -0.10$, did not show a difference for the correct first identification for experienced participants ($M = 3.92, SD = 0.29$) or expert participants ($M = 3.89, SD = 0.33$). Detection, $F(1,20) =$

$28.96, p < .01, d = 2.38$, had fewer occurrences of correct first identification for experienced participants ($M = 1, SD = 1.35$) than for expert participants ($M = 3.67, SD = 0.71$). Response implementation, $F(1,20) = 0.04, p = .84, d = -0.10$, did not show a difference for the correct first identifications for experienced participants ($M = 3.92, SD = 0.29$) or expert participants ($M = 3.89, SD = 0.33$).

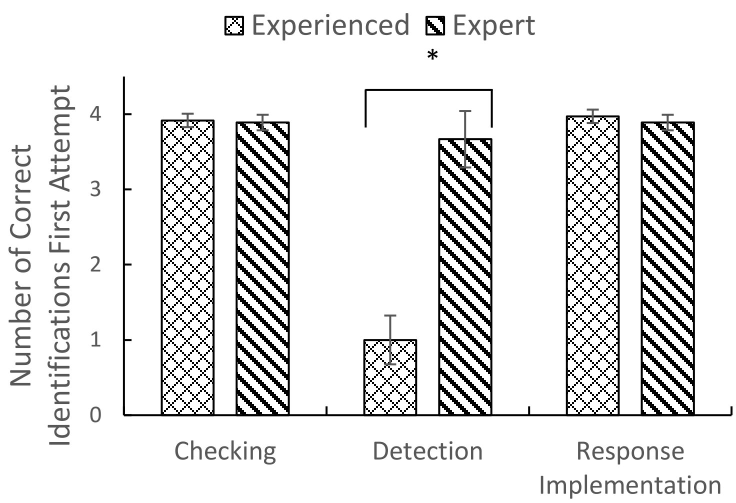

Figure 43: RO2 number of correct identifications on first attempt by task type and experience level. Each task type has a potential of four identifications. Error bars represent standard error. Asterisk represents statistical significance.

For the number of additional identifications made on the correct I&C, a main effect was found for task type, $F(1.05,19.93) = 31.39, p < .01, \eta_p{}^2 = .62$, such that detection ($M = 6.33, SD = 9.20$) was higher than both checking ($M = 0.33, SD = 0.91$) and response implementation ($M = 0.19, SD = 0.40$). A main effect for experience level was found for the number of additional identifications made on the correct I&C, $F(1,19) = 31.11, p < .01, \eta_p{}^2 = .62$, such that experienced participants ($M = 0.17, SD = 0.22$) performed fewer additional actions compared to expert participants ($M = 5.11, SD = 3.09$). A significant interaction effect was found between task type and experience level for the number of additional times a correct I&C identification action was performed, $F(1.05, 19.93) = 31.39, p < .01, {\eta_p}^2 = .62$, (Figure 44). Checking, $F(1,20) = 0.93, p = .35, d = 0.23$, had no difference in additional actions for experienced participants ($M = 0.17, SD = 0.39$) or expert participants ($M = 0.56, SD = 1.33$). Detection, $F(1,23.80) = 8.05, p < .01, d = 0.03$, had fewer additional actions for experienced participants ($M = 0.17, SD = 0.58$) than for expert participants ($M = 14.56, SD = 8.83$). Response implementation, $F(1,20) = 0.09, p = .76, d = 0.21$, had no difference in additional actions between experienced participants ($M = 0.17, SD = 0.39$) and expert participants ($M = 0.22, SD = 0.44$).

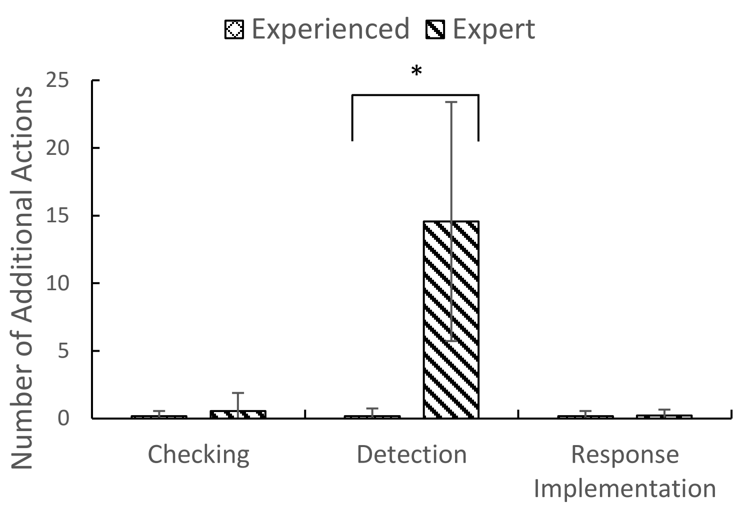

Figure 44: RO2 number of additional identifications on the correct I&C by task type and experience level. Error bars represent standard error. Asterisk represents statistical significance.

For the number of identifications made on I&Cs not relevant to the task, no main effect was found for task type, ($p = .07$). A main effect for experience level was found for the number of identifications made on I&Cs not relevant to the task, $F(1,19) = 10.64, p < .01, {\eta_p}^2 = .36$, such that experienced participants ($M = 0, SD = 0$) had fewer erroneous actions compared to expert participants ($M = 0.26, SD = 0.28$). No interaction effect was found between the task type and experience level for the number of identifications made on I&Cs not relevant to the task, ($p = .07$).

Action

Independent sample t-tests were conducted to determine if there were significant differences between experienced and expert participants for various action performance variables. Below are the descriptions and results of each action performance measure for the detection and response implementation task types.

Detection

The percentage of correct gauge change detections, percentage of missed gauge change events, and the number of false positive detections for each participant were measured while completing the detection task type.

A significant difference was found for the percentage of correct gauge change detections; such that, experienced participants ($M = 85.10, SD = 6.48$) had a greater percentage of correctly detected gauge change events than expert participants ($M = 33.21, SD = 23.51$); $\mathrm{t}(8.92) = 6.44, \mathrm{p} < .01$ (Figure 45).

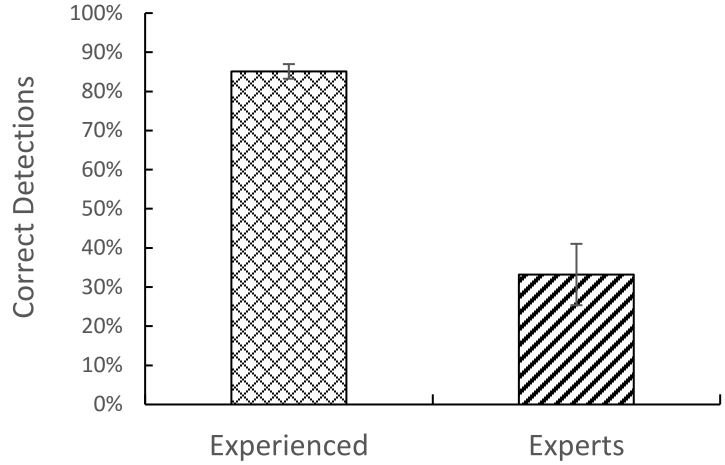

Figure 45: RO2 percentages of correctly acknowledged gauge detections by experience level. Error bars represent standard error.

A significant difference was found for percentage of missed gauge change events; such that, experienced participants ($M = 11.98, SD = 5.03$) had a lower percentage of missed gauge change events than expert participants ($M = 58.61, SD = 26.55$); $\mathrm{t}(8.43) = -5.20, \mathrm{p} < .01$ (Figure 46).

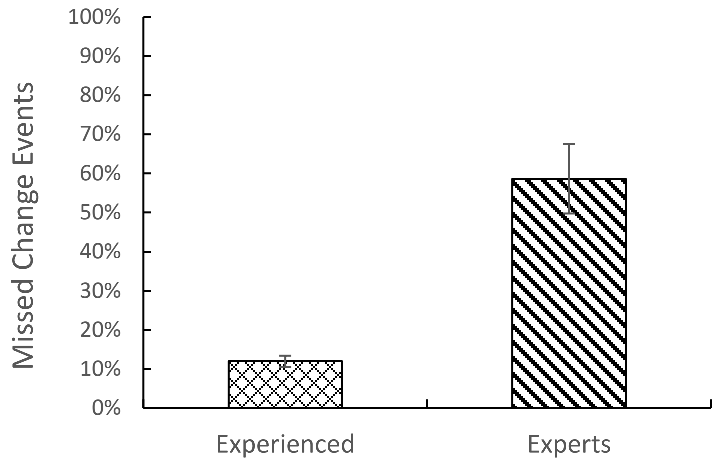

Figure 46: RO2 percentages of missed gauge change events by experience level. Error bars represent standard error.

A significant difference was found for number of false positive detections; such that, experienced participants ($M = 8.58, SD = 5.14$) had a fewer false positive detections than expert participants ($M = 78.78, SD = 42.32$); $\mathrm{t}(8.18) = -4.95, \mathrm{p} < .01$ (Figure 47).

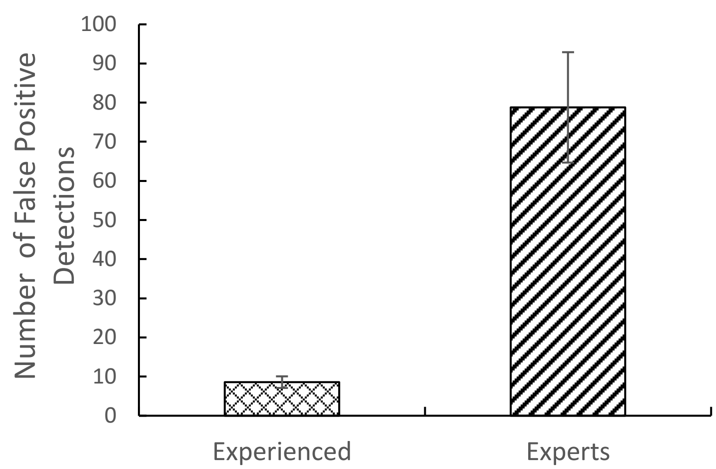

Figure 47: RO2 number of false positive acknowledgements by experience level. Error bars represent standard error.

Response Implementation

The number of times a participant followed the correct sequence of identifying the I&C and then manipulating it in the correct direction, the percent of percentage of correct manipulations, percentage of description error, and the percentage of mode errors for each participant was measured while completing the response implementation task type.

No differences in the number of correctly followed sequences was found for experienced participants ($M = 0.58, SD = 0.52$) and expert participants ($M = 0.56, SD = 0.56$); (p = .91). No difference in correct manipulations was found between experienced participants ($M = 81.79, SD = 26.43$) and expert participants ($M = 82.22, SD = 22.79$); (p = .97). No differences in the number of description error manipulations was found for experienced participants ($M = 1.66, SD = 1.66$) and expert participants ($M = 3.70, SD = 3.70$); (p = .59). No difference was found for mode error manipulations for experienced participants ($M = 7.34, SD = 14.41$) and expert participants ($M = 4.07, SD = 2.71$); (p = .55).

Physiological Measures

All dependent variables that were used in the ANOVAs were calculated by taking a difference from a five-minute wakeful resting baseline. For example, if the participant's heart rate HR for the five-minute baseline was 53 beats per minute (BPM) and their HR for the subsequent detection task was 62 BPM, their difference from baseline would be 9 BPM, ($62 - 53 = 9$). This approach helps account for individual differences when comparing group means in an ANOVA.

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) as well as lobes (i.e., compare brain activity among the frontal, parietal and occipital lobes). Hence, a series of ANOVAs were performed.

Hemispheres

Six mixed 3 (*Task Type*: checking, detection, and response implementation) × 2 (*Experience Level*: experienced and expert) ANOVAs were run for left and right hemispheres for theta, alpha, and beta frequency bands. These ANOVAs provided insight in to the overall effects of task type and experience level on the left and right hemispheres.

Theta in the Left and Right Hemispheres

For theta in the left hemisphere, a main effect was found for task type, $F(1.34, 25.44) = 12.77, p < .01, {\eta_p}^2 = .40$, such that detection ($M = -62.70, SD = 572.81$) had less of a change from baseline than both checking ($M = 797.53, SD = 1240.37$) and response implementation ($M = 820.66, SD = 1526.33$).

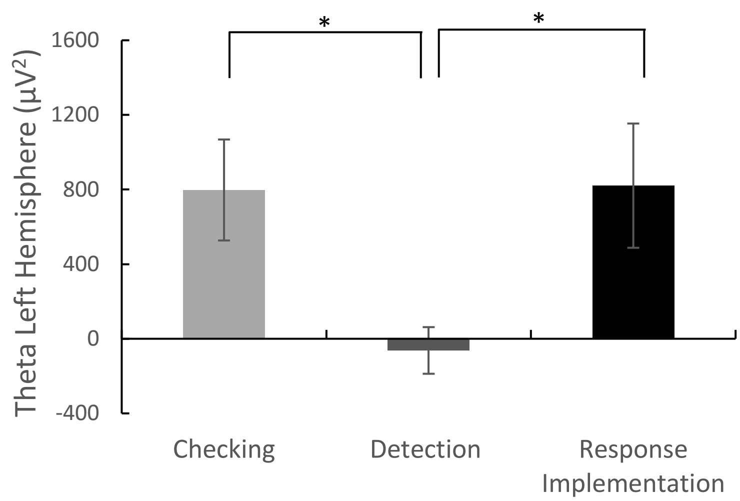

Figure 48: RO2 theta left hemisphere difference from baseline in μV^2 for the three task types. Error bars represent standard error. Asterisk represents statistical significance.

Also, a main effect was found for experience level for theta in the left hemisphere, $F(1,19) = 7.60, p = .01, {\eta_p}^2 = .29$, (Figure 49) such that experienced participants ($M = 52.52, SD = 370.06$) had lower theta in the left hemisphere compared to expert participants ($M = 1139.81, SD = 1308.14$).

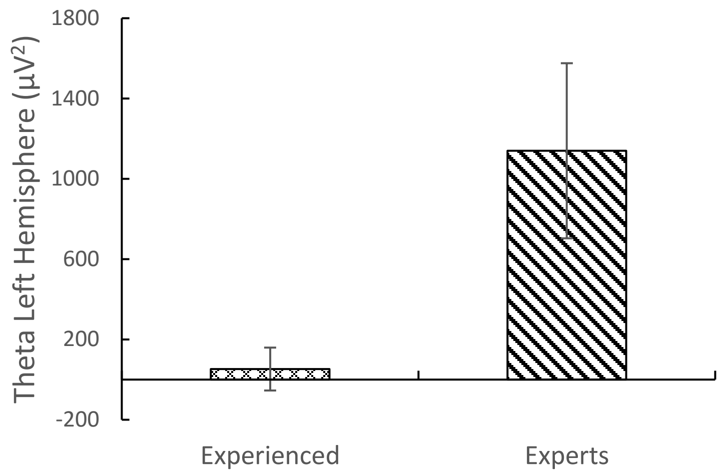

Figure 49: RO2 theta left hemisphere difference from baseline in μV² by experience level. Error bars represent standard error.

No interaction effect was found between task type and experience level for theta in the left hemisphere, $(p = .11)$.

For theta in the right hemisphere, a main effect was found for task type, $F(2{,}38) = 33.43, p < .01, {\eta_p}^2 = .64$, (Figure 50)such that detection $(M = -32.63, SD = 545.79)$ had less of a change from baseline than both checking $(M = 561.24, SD = 670.95)$ and response implementation $(M = 545.99, SD = 745.30)$.

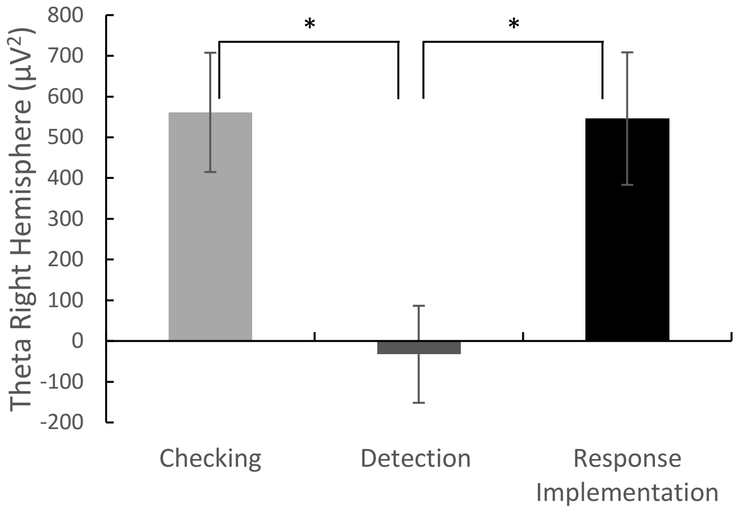

Figure 50: RO2 theta right hemisphere difference from baseline in μV^2 for the three task types. Error bars represent standard error. Asterisk represents statistical significance.

A main effect was found for experience level for theta in the right hemisphere, $F(1,19) = 13.14, p < .01, {\eta_p}^2 = .41$, (Figure 51) such that experienced participants ($M = 21.99, SD = 352.10$) had lower theta in the right hemisphere compared to expert participants ($M = 806.49, SD = 633.85$).

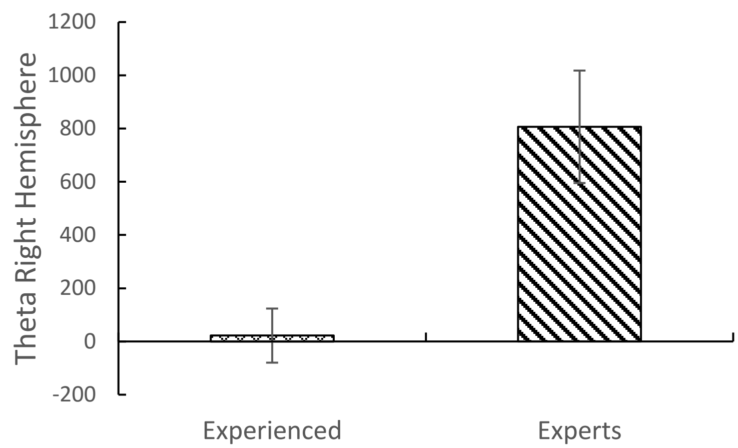

Figure 51: RO2 theta right hemisphere difference from baseline in μV^2 by experience level. Error bars represent standard error.

No interaction effect was found between task type and experience level for theta in the right hemisphere, ($p = .62$).

Alpha in the Left and Right Hemispheres

For alpha in the left hemisphere, no main effect was found for task type, ($p = .08$), or experience level, ($p = .60$). Also, no interaction effect was found between task type and experience level for alpha in the left hemisphere, ($p = .11$).

For alpha in the right hemisphere, a main effect was found for task type, $F(1.31,24.81) = 3.92, p = .05, {\eta_p}^2 = .17$. However, pairwise comparisons did not show any significant differences between the checking ($M = 181.23, SD = 1895.31$), detection ($M = -288, SD = 1011.69$), and response implementation ($M = 332.39, SD = 2153.72$). No main effect for experience level was found for alpha in the right hemisphere, ($p = .13$). No interaction effect was found between task type and experience level for alpha in the right hemisphere, ($p =$ .07).

Beta in the Left and Right Hemispheres

For beta in the left hemisphere, a main effect was found for task type, $F(2,38) = 10.14, p < .01, {\eta_p}^2 = .35$, (Figure 52) such that detection ($M = 456.50, SD = 1532.79$) had less of a change from baseline than both checking ($M = 1996.64, SD = 1415.84$) and response implementation ($M = 1856.55, SD = 760.60$).

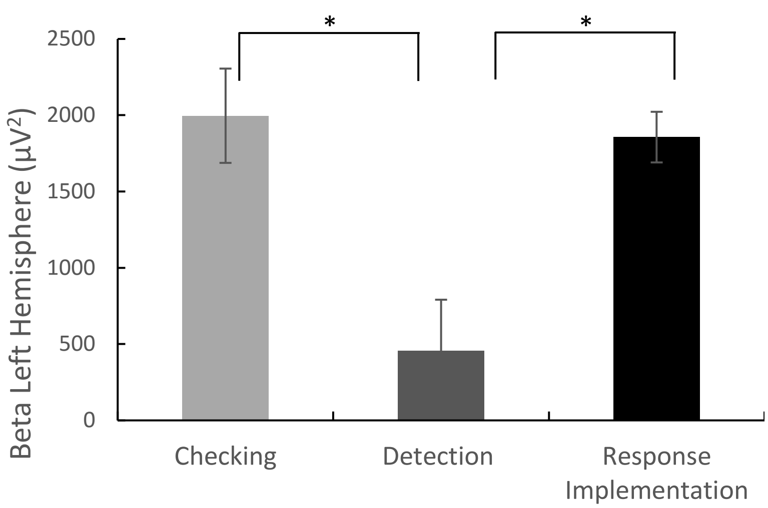

Figure 52: RO2 beta left hemisphere difference from baseline in μV^2 for the three task types. Error bars represent standard error. Asterisk represents statistical significance.

A main effect was found for experience level for beta in the left hemisphere, $F(1,19) = 4.68, p = .04, {\eta_p}^2 = .20$, (Figure 53) such that experienced participants ($M = 1135.99, SD = 377.50$) had lower beta in the left hemisphere compared to expert participants ($M = 1837.33, SD = 1043.68$).

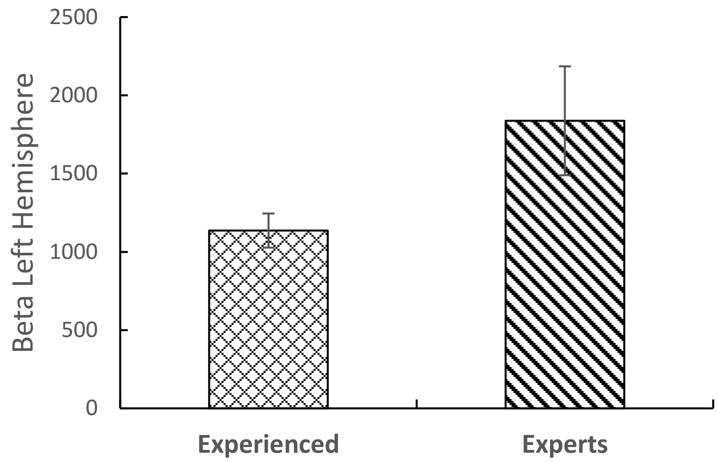

Figure 53: RO2 beta left hemisphere difference from baseline in μV^2 by experience level. Error bars represent standard error.

No interaction effect was found between task type and experience level for beta in the left hemisphere, $(p = .32)$.

For beta in the right hemisphere, a main effect was found for task type, $F(2,38) = 25.33, p < .01, {\eta_p}^2 = .57$, (Figure 54) such that detection $(M = 277.03, SD = 1247.37)$ had less of a change from baseline than both checking $(M = 2113.00, SD = 1506.63)$ and response implementation $(M = 2454.43, SD = 1322.26)$.

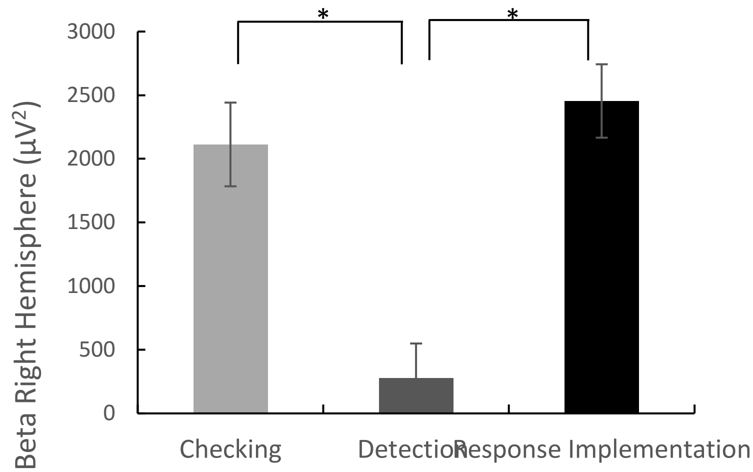

Figure 54: RO2 beta right hemisphere difference from baseline in μV^2 for the three task types. Error bars represent standard error. Asterisk represents statistical significance.

No main effect for experience level was found for beta in the right hemisphere, $(p = .83)$. No interaction effect was found between task type and experience level for beta in the right hemisphere, $(p = .60)$.

Lobes

Nine mixed 3 (*Task Type*: checking, detection, and response implementation) × 2 (*Experience Level*: novice and expert) ANOVAs were run for frontal, parietal and occipital lobes for theta, alpha, and beta frequency bands. These ANOVAs provided insight in to the overall effects of task type and experience level on the frontal, parietal and occipital lobes.Theta in the Frontal, Parietal, and Occipital Lobes

For theta in the frontal lobe, a main effect was found for task type, $F(1.15, 21.91) = 9.01, p < .01, {\eta_p}^2 = .32$, (Figure 55) such that detection $(M = -52.76, SD = 665.73)$ had less of a change from baseline than both checking ($M = 650.74, SD = 1279.70$) and response implementation ($M = 688.57, SD = 1562.30$).

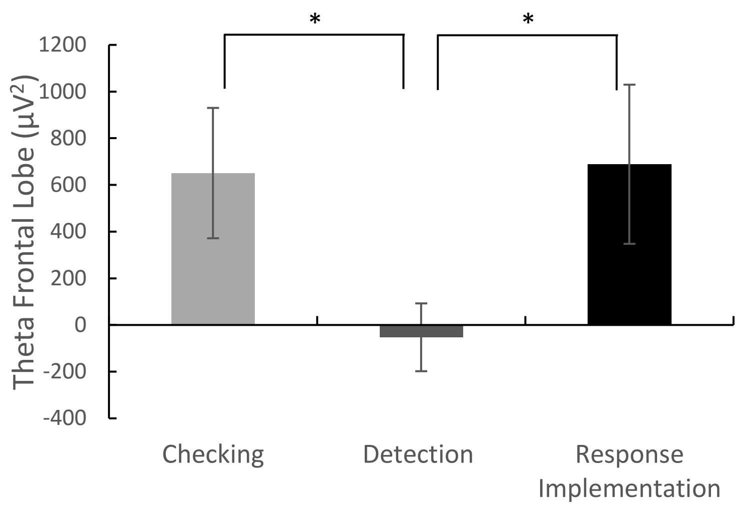

Figure 55: RO2 theta frontal lobe difference from baseline in μV^2 for the three task types. Error bars represent standard error. Asterisk represents statistical significance.

A main effect was found for experience level for theta in the frontal lobe, $F(1,19) = 9.19, p < .01, \eta_p{}^2 = .33$, (Figure 56) such that experienced participants ($M = -96.79, SD = 395.63$) had lower theta in the frontal lobe compared to expert participants ($M = 1129.70, SD = 1335.75$).

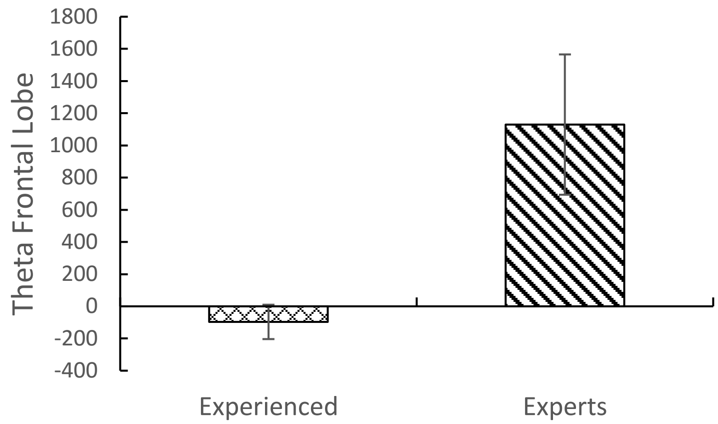

Figure 56: RO2 theta frontal lobe difference from baseline in μV^2 by experience level. Error bars represent standard error.

No interaction effect was found between task type and experience level for theta in the frontal lobe, ($p = .17$).

For theta in the parietal lobe, a main effect was found for task type, $F(2{,}38) = 43.00, p < .01, {\eta_p}^2 = .69$, (Figure 57) such that detection ($M = -174.43, SD = 619.85$) had less of a change from baseline than both checking ($M = 740.35, SD = 722.12$) and response implementation ($M = 624.35, SD = 579.89$).

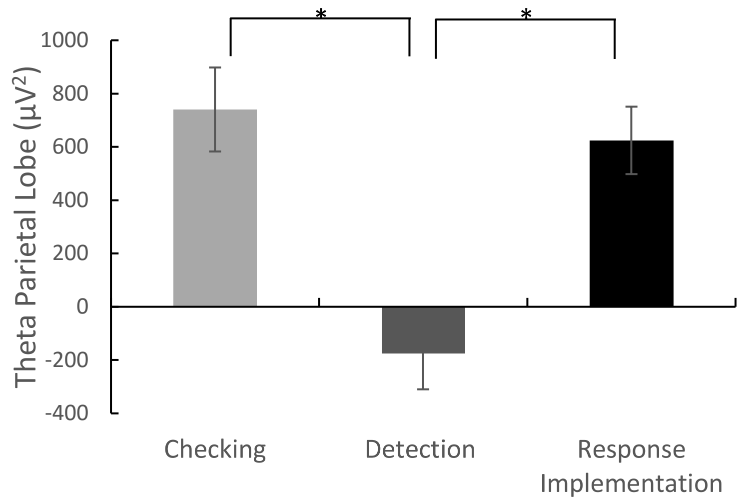

Figure 57: RO2 theta parietal lobe difference from baseline in μV^2 for the three task types. Error bars represent standard error. Asterisk represents statistical significance.

A main effect was found for experience level for theta in the parietal lobe, $F(1,19) = 10.93, p < .01, {\eta_p}^2 = .37$, (Figure 58) such that experienced participants ($M = 101.39, SD = 368.24$) had lower theta in the parietal lobe compared to expert participants($M = 790.59, SD = 587.04$).

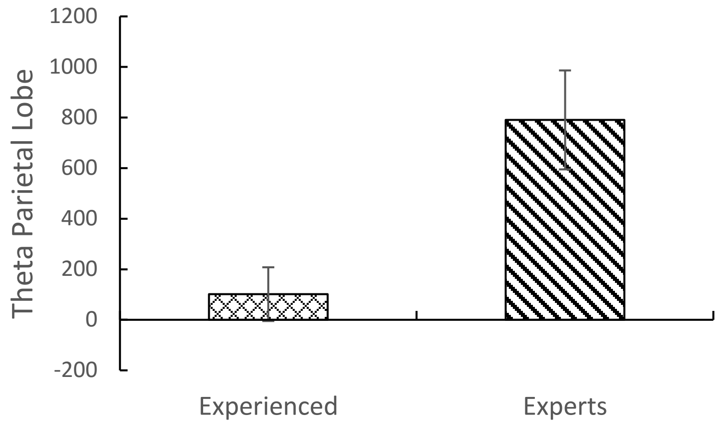

Figure 58: RO2 theta parietal lobe difference from baseline in μV^2 by experience level. Error bars represent standard error.

No interaction effect was found between task type and experience level for theta in the parietal lobe, ($p = .49$).

For theta in the occipital lobe, no main effect was found for task type, ($p = .34$). A main effect was found for experience level for theta in the occipital lobe, $F(1,19) = 4.50, p = .05, {\eta_p}^2 = .19$, (Figure 59) such that experienced participants ($M = 55.45, SD = 211.85$) had lower theta in the occipital lobe compared to expert participants ($M = 352.00, SD = 420.83$).

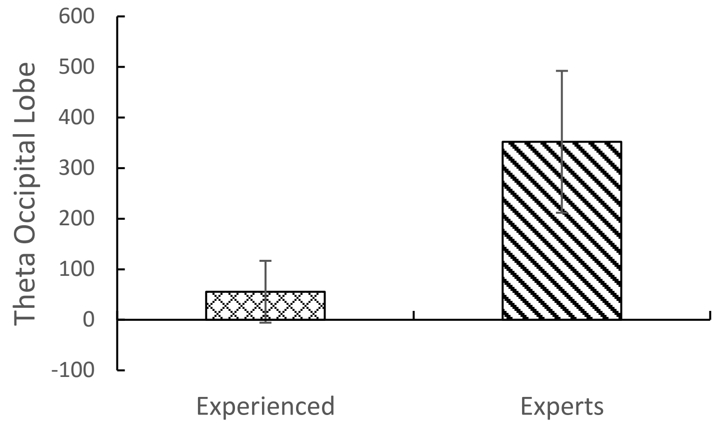

Figure 59: RO2 theta occipital lobe difference from baseline in μV^2 by experience level. Error bars represent standard error.

No interaction effect was found between task type and experience level for theta in the occipital lobe, ($p = .35$).

Alpha in the Frontal, Parietal, and Occipital Lobes

For alpha in the frontal lobe, no main effect was found for task type, ($p = .25$) or experience level, ($p = .50$). No interaction effect was found between task type and experience level for alpha in the frontal lobe, ($p = .07$).

For alpha in the parietal lobe, a main effect was found for task type, $F(2,38) = 6.28, p < .01, {\eta_p}^2 = .25$, (Figure 60) such that detection ($M = -469.08, SD = 964.52$) had a decrease from baseline which differed from checking ($M = 116.18, SD = 1265.15$), and response implementation ($M = 32.09, SD = 914.35$) did not differ from either checking or detection.

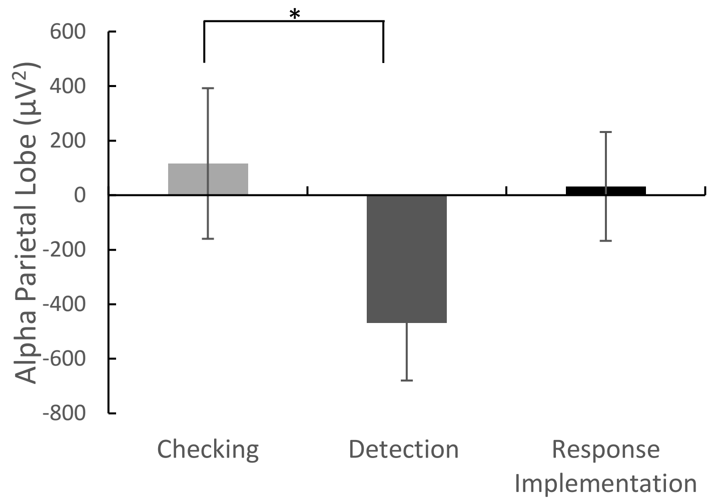

Figure 60: RO2 alpha parietal lobe difference from baseline in μV^2 for the three task types. Error bars represent standard error. Asterisk represents statistical significance.

No main effect was found for experience level for alpha in the parietal lobe, ($p = .50$). No interaction effect was found between task type and experience level for alpha in the parietal lobe, ($p = .15$).

For alpha in the occipital lobe, no main effect was found for task type, ($p = .23$), or experience level, ($p = .41$). Also, no interaction effect was found between the task type and experience level for alpha in the occipital lobe, ($p = .15$).

Beta in the Frontal, Parietal, and Occipital Lobes

For beta in the frontal lobe, a main effect was found for task type, $F(2,38) = 17.12, p < .01, {\eta_p}^2 = .47$, (Figure 61) such that detection ($M = 296.56, SD = 1274.16$) had less of a change from baseline than checking ($M = 1897.22, SD = 1102.08$), and response implementation ($M = 1974.23, SD = 848.20$) did not differ from either checking or detection.

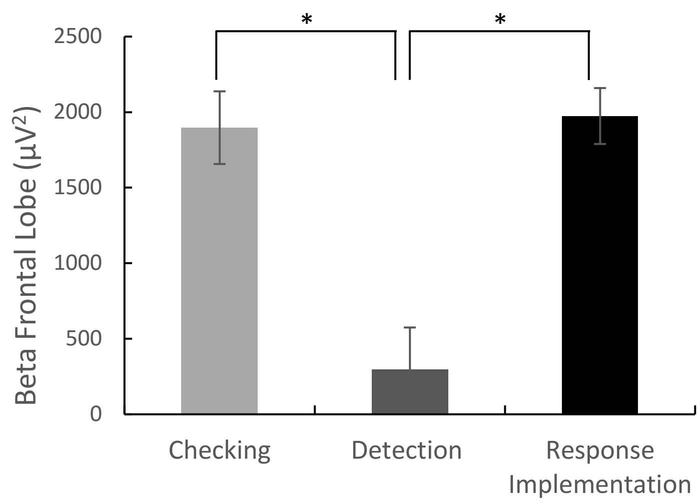

Figure 61: RO2 beta frontal lobe difference from baseline in μV^2 for the three task types. Error bars represent standard error. Asterisk represents statistical significance.

No main effect was found for experience level for beta in the frontal lobe, ($p = .41$). No interaction effect was found between task type and experience level for beta in the frontal lobe, ($p = .32$).

For beta in the parietal lobe, a main effect was found for task type, $F(2,38) = 15.02, p < .01, {\eta_p}^2 = .44$, (Figure 62) such that detection ($M = 37.43, SD = 1441.51$) had less of a change from baseline than checking ($M = 1910.58, SD = 1077.87$), and response implementation ($M = 1771.04, SD = 1184.32$) did not differ from either checking or detection.

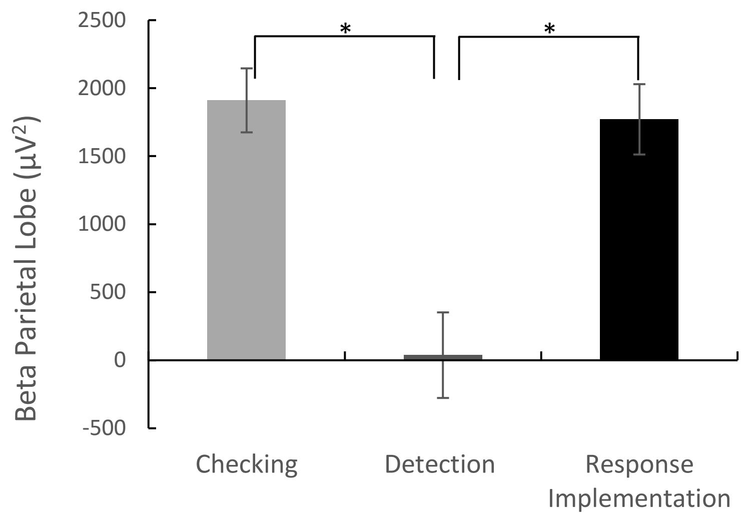

Figure 62: RO2 beta parietal lobe difference from baseline in μV^2 for the three task types. Error bars represent standard error. Asterisk represents statistical significance.

There was no main effect for experience level for beta in the parietal lobe, ($p = .56$). No interaction effect was found between task type and experience level for beta in the parietal lobe, ($p = .32$).

For beta in the occipital lobe, a main effect was found for task type, $F(2,38) = 10.55, p < .01, {\eta_p}^2 = .36$, (Figure 63) such that detection ($M = 295.67, SD = 1224.22$) had less of a change from baseline than both checking ($M = 1192.08, SD = 1112.34$) and response implementation ($M = 1601.21, SD = 1785.11$).

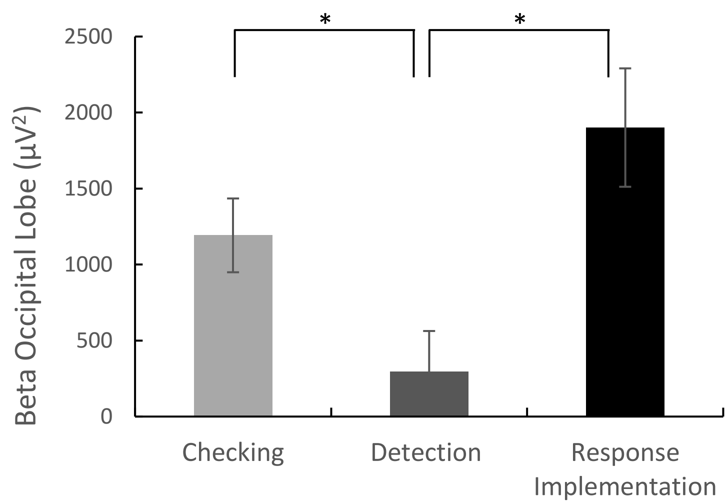

Figure 63: RO2 beta occipital lobe difference from baseline in μV^2 for the three task types. Error bars represent standard error. Asterisk represents statistical significance.

No main effect was found for experience level for beta in the occipital lobe, ($p = .99$). No interaction effect was found between task type and experience level for beta in the occipital lobe, ($p = .59$).

Electrocardiogram (ECG)

Three mixed 3 (*Task Type*: checking, detection, and response implementation) × 2 (*Experience Level*: experienced and expert) ANOVAs were conducted to determine if the different task types and experience levels impacted HR, HRV, and/or IBI. These analyses also assessed the interactive effects between the task types and experience level, which would reveal if any of the observed differences across task types were similar for experienced and expert participants. Task type was a repeated-measures variable and experience level was a between-participants variable. All measures were calculated as a difference from baseline to mitigate individual differences.

Heart rate (HR)

HR was derived from R-Peak detections using the So-Chan QRS algorithm from a raw electrical ECG signal. A main effect was found for task type for HR, $F(1.25,23.79) = 4.59, p = .04, {\eta_p}^2 = .19$, (Figure 64). However, pairwise comparisons did not show any difference between checking ($M = -6.88, SD = 12.11$), detection ($M = -16.37, SD = 26.19$), and response implementation ($M = -6.83, SD = 10.81$). No main effect was found for experience level for HR, ($p = .83$). No interaction effect was found between task type and experience level for HR, ($p = .42$).

Interbeat Interval (IBI)

IBI was derived from R-Peak detections using the So-Chan QRS algorithm from a raw ECG electrical signal. For IBI, no main effect was found for task type, ($p = .92$), or experience level, ($p = .21$). No interaction effect was found between task type and experience level for IBI, ($p = .38$).

Heart Rate Variability (HRV)

For HRV, no main effect was found for task type, ($p = .33$), or experience level, ($p = .11$). No interaction effect was found between task type and experience level for HRV, ($p = .20$).

DISCUSSION

The analysis revealed several interesting findings for workload levels between the three task types studied (checking, detection, and response implementation) that were persistent across the different levels of experience. It's worth noting that while several measures had statistical significance between checking, detection, and response implementation task types, there was no single measure that was sufficiently effective at distinguishing between each of the three task types.

In addition to the task type findings, the various levels of experience also revealed interesting findings for workload. All of the findings that relate to the aims of this research will be discussed in more details below. The discussion section will be structured in the following way. First the results will be explained for the novice-expert comparison group and then for experienced-expert comparison group. Within each comparison group, the meaningful finding will be discussed by (i) perceived workload, (ii) performance measures, and (iii) physiological response to task demands. Lastly, the discussion section concludes with the research findings and their implications for the research community interested in workload studies in the NPP domain.

The goal for this research is to validate the "equal but different" principle for using novices as a proxy for expert ROs when investigating levels and types of workload during NPP MCR operations. The aim of this research is to determine if certain workload measures are sufficiently effective across skill level at assessing types of workload experienced during NPP MCR task types. In addition to types of workload experienced, this research aims to measure trends in terms of workload level across the experience continuum form novice to expert and then from experience to expert ROs.

Novice-Expert Comparison

Perceived Workload

The detection task type, in general was rated the highest for workload for both novices and experts as reported from surveys taken immediately upon completing each of the task types. This finding is consistent with Mercado's study that looked at workload across these same three task types (Joseph E Mercado, Lauren Reinerman-Jones, Daniel Barber, & Rebecca Leis, 2014). Physical demand and frustration are the two subscale types from the NASA TLX that were the biggest driver of the workload differences between each of the task types. It's not surprising that these two types of workload are higher for the detection task type. The detection task type differs from the other two task types in several ways. First, the detection task type took much longer to complete. It lasted on average 24 minutes and 41 seconds whereas the other two task types took on average 4 minutes and 25 seconds to complete. This is because the detection included four sustained attention steps that lasted five minutes each. Within each of these four steps, the participants were required to continuously monitor a single gauge and report every discrete change in the gauge's value by pressing the gauge label directly below the gauge. The detection task had 240 discrete changes that occurred on the four gauges that needed to be reported. In addition to the sustained attention required for acknowledging gauge changes, participants were also required to keep in working memory the gauge's threshold value at which they needed to report back to the SRO once it crosses. These differences in the detection task type compared to the checking and response implementation task types are what drive the major changes in levels and types of workload.

The experts overall across the three task types tended to rate level of perceived workload lower than novices. This is consistent with intuition. Experts have many more years of experience performing these task types so naturally workload should be lower.

Although there were statistical differences for the levels of perceived workload ratings between novices and experts, the same overall types of workload (as reported in the NASA TLX) were seen regardless of experience level.

Performance

In general, the performance measures ware consistent across the three task types. There was one noticeable exception to this finding. The number of instructions that were requested to be repeated was much higher for the detection task type. This is consistent with the TLX frustration and performance ratings for the detection task. Instructions for the checking task type require the participant to verify a specific valve or light box is tripped, open, or shut. Instructions for the response implementation task type require the participant to verify then open or shut a specific valve. Instructions for the detection task type require the participant to verify then report back once a specific gauge crosses a threshold value. The difference for the detection task type instruction is that participants were required to remember two numbers. This is because the gauge name contains a number as does the gauge report back threshold. It was observed that the repeat back of the initial SRO instruction would often fail at the threshold value before a participant would request a repeat instruction (e.g., "SRO, understood you want me to verify LI 494 Sierra Alpha and report when…SRO, Pease repeat"). Also, the experts (although not part of experimental training) would often request just the report back threshold (e.g., “SRO please repeat the threshold for LI 494 Sierra Alpha”).

Experts were either no different or performed better than the novices for all the performance measures with the exception of one notable measure. Novices had significantly higher percentage of communications completed correctly. Upon manual review of the transcription logs, a common cause for experts having incorrect communications was uncovered. Experts often left off the addressee in their communication when it was implied they were speaking to the SRO. However, strict scoring of three-way communications for the experiment required an addressee/recipient be at the start of each communication. The reason for experts underperforming when it comes to communications completed correctly is negative transfer effects from plant specific training. Three-way communications, while not mandated by regulation, is practiced and trained at most NPPs. The lack of a common standard has led to plants implementing different variants of three-way communications. Some plants do not require an addressee/recipient when it’s implied as in the case of a reply to an instruction. It was observed that experts tended to have negative transfer effects when trying to perform to the strict standard used in the experiment. The biggest issue by far for experts was not addressing the recipient of the communication (e.g., an expert would say “that is correct” rather than saying “SRO, that is correct”).

Experts tended to outperform novices in identifying correct I&Cs and performed far fewer mode errors during the response implementation task type. These measures are important to experts because committing these types of errors are costly. During a license evaluation, committing one of these errors could result in failure to get your license renewed.

While there was a significant difference in several performance measures, such that experts performed better than novices, novices still performed the majority of tasks correctly. However, that not to say that novice performance measures in general aren't valuable at assessing workload. Novices tend to commit the same types of errors as experts just not at the same rates.

Physiological Response

Overall, theta and beta in several regions of the brain showed less change from baseline for detection than both checking and response implementation. 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. However, vigilance tasks have shown that over time, attention cannot be sustained and is paralleled by a decrement in performance and physiological response (Berka et al., 2007b). Additionally, the detection task is longer than both checking and response implementation but has the same number of navigation and identification steps. While the working memory component of detection is harder (as indicated by performance) than the other task types, the portion of time during the task where the participant had to hold the gauge name in working memory is less than the other task types. The working memory and concentration task demands are seen in the increase in theta across both hemispheres for the all three task types. These demands are indicated by the fact that theta increased from the resting baseline. This is consistent with the fact that all three task types had the navigation and identification component which among other demands also required participants to hold information in working memory. Theta has a larger increase from baseline for both checking and response implementation compared to detection. Detection’s overall portion of task time for navigation and identification was significantly less than both checking and response implementation. Theta increase and specifically the larger increase for checking and response implementation, shows that navigation and identification task components were the major source of working memory and concentration demands.

Beta for all three task types showed an increased from baseline. This increase is a direct reflection of the cognitive processing demands imposed by the tasks. However, beta increase from baseline for detection was less than both checking and response implementation. Beta is a reflection of the participant’s arousal and attention. The fact that detection had less of an increase than both checking and response implementation, shows that the sustained attention component when monitoring a gauge acts like a vigilance task. Vigilance tasks are often associated with a drop in attention and arousal. Detection lasted around 24 minutes which shows how quickly vigilance tasks can impact attention. Perceived global workload ratings were higher for detection compared to checking and response implementation, which provides further support for this task type reflecting a vigilance task.

Cognitive processing has been shown to influence HRV, such that more mental processing results in a lower HRV (Luque-Casado, Zabala, Morales, Mateo-March, & Sanabria, 2013). Detection had a lower HRV than the response implementation task type. A lower HRV is associated with an increase in workload (L. J. Mulder, D. de Waard, & K. A. Brookhuis, 2004). Perceived workload, primary task performance, and physiological correlates were all in agreement that detection had the largest workload.

In general, novices performing the modified "equal but different" EOP with the reduced instructions and simplified panel had equivalent physiological workload responses as the experts that performed the modified EOP with the full instruction and fully populated panels.

Experienced-Expert Comparison

Perceived Workload

The detection task type, in general was rated the highest for workload by both experienced and expert participants as reported from surveys taken immediately upon completion of each task type. Like the novice-expert comparison, experienced and experts found physical demand to be a significant driver of workload. Unlike the novice-expert comparison, frustration was not as big of an overall factor for the detection task type workload.

The experienced participants across all of the perceived workload measure recorded in this study reported significantly lower ratings than experts. Further investigation into the experienced ratings revealed that they reported very low ratings across every scale. This is likely due to the fact that they repeated the exact experiment multiple times. After several identical runs of the experiment, tasking became second nature for the experienced participants. Over time the perceived measures of workload became lower and lower.

Performance

The performance measures are in agreement with perceived workload ratings. Experienced participants reported lower workload ratings and had better performance in general compared to the experts. The experienced participants outperformed the experts in all performance measure with the exception of identifying the correct I&C. Further investigation showed that for checking and response implementation, both experienced and expert identified 100% of the correct controls; however, for the detection task type experienced participants did not identifying the gauge correctly. Reviewing the log files generated from the simulator revealed that experienced participants were incorrectly identifying gauges by pressing the gauge label button rather than touching the gauge itself. Training for the experiment specifically covered identification as touching directly on I&Cs and not the gauge label button. This is a human factor issue of the interface seen with both novice and experienced participants. Pressing an adjacent button associated with an I&C appears to be a natural way of identifying the gauge. Experts did not have this issue because they have been trained for years that touching the I&C is how to signal you have identified the proper I&C. This interface identification issue was only present in the detection task type because the other I&Cs did not have associated buttons.

Physiological Response

Alpha recorded in the parietal lobe when averaged over the task type were lower for detection compared to checking and response implementation. A suppressed alpha is correlated with increases in workload (Hankins & Wilson, 1998). Specifically, alpha is believed to be suppressed during visually tasks due to retinal and oculomotor control. This is see mostly in the parietal and occipital lobes (Wertheim, 1981). Detection require participants to spend a significant portion of the task time in constant focus on gauges.

Theta and beta increased from the resting baseline for the three task types across all regions of the brain. This is consistent with the novice-expert comparisons as well as in Mercado’s (2014) study which used the same three task type. This shows that the tasks demands for the task types are increasing working memory, concentration, attention and arousal demands regardless of experience level. Also consistent with the novice-expert participants, theta and beta were lower for detection compared with checking and response implementation.

The experienced participants had a lower theta in multiple regions of the brain than the experts regardless of task type. Repeated trials of an experiment induces lower workload as indicated by the suppressed theta for the experienced participants. Theta is correlated most strongly with working memory and concentration (Kubota et al., 2001). By repeating trials, working memory and concentration task demands are lessened. Perception, performance, and physiological response are all in agreement. Experienced participants due actually experience lower workload than the expert participants in this experiment.

Conclusion

The goal for this research was to validate the "equal but different" principle for using novices as a proxy for expert ROs when investigating levels and types of workload during NPP MCR operations. This research evaluated novices and experts by looking at level and types of workload through questionnaires, performance measures and physiological correlates when performing three task types (checking, detection and response implementation). In addition to the novice-expert groups, comparisons were also made against experienced-expert groups. This allowed for trend analysis as skill level is increased across the Dreyfus continuum.

While there were observable differences in the perceived levels of workload between novice and experts, such that experts perceived slightly lower workload. The main differences in workload sources, physical demand and frustration, for detection compared to checking and detection were consistent between the novices and experts. However, the anticipated workload trend (novice → experienced → expert) across the skill continuum was not observed. This is due to the repeated experimental sessions for only the experienced participants. Repeated sessions reduces several factors that influence workload and is observed through perceived workload, performance measures, and physiological workload correlates. The "equal but different" principal (novice ≈ expert) for the NPP MCR tasks was found to induce equal cognitive demands between novices and experts. The parallels in workload perception, performance measures, and physiological responses between novices and experts were consistent across checking, detection, and response implementation task types. While novices and experts were not identical in all measures of workload, there were meaningful similarities between the two experience levels for workload.

Performance measures for workload were more similar between novices and experts than for experienced and experts. These similarities were seen across the three task types. Experienced participants outperformed both novices and experts on most tasks. However, experts did perform the best of all groups on the reactor license critical task of identifying the correct I&C and opening/closing the desired I&C correctly. Performing incorrectly on these tasks during an evaluation could result in not having their license renewed. Correctly performing three-way communications is where experts had the most trouble and actually performed worse than novices. Specifically, the common issue among some of the experts was omitting the recipient identifier in their communications. Some of the operators reported that their plant didn't mandate the recipient if it was implied in a response. Besides the mission critical tasks (performance: novices < experts) and the three-way communication issues (performance: novices > experts), novices using the reduced verbal instruction and simplified I&C panels performed the modified EOP on par with the experts using the full instruction and fully simulated I&C panel.

On all performance measures (with the exception of correct I&C identification) experienced participants performed at least as well as if not better than novices and experts. This again goes against the anticipated trend of performance expectations that performance should increase (novice < experienced < expert) across the skill continuum. However, this finding shows that experts not operating on their home plant are disadvantaged when it comes to plant specific practices.

The simulator used in this experiment was a replica of the Shearon Harris NPP and the SROs verbal three-way communication instructions were derived from that plants procedures. One expert participant had operational experience at the Shearon Harris plant. That participant identified 100% of the correct I&Cs on the first attempt and also did not perform any invalid actions during all three task types. While the home plant expert is only one of the experts, it appears that experts perform better with a plant layout and procedures they are familiar with.

Physiological response of the brain as recorded by theta, alpha, and beta showed no difference between novices with the reduced instructions and simplified I&C panels and experts with the complete instruction set and full I&C panel. This finding supports the use of the “equal but different” principle because both groups while not different from each other, were sensitive to workload measures from the three different task types. Specifically, the theta findings show that regardless of experience for novice or expert populations, the checking and response implementation tasks had a larger percentage of their time requiring working memory and concentration compared to detection. Beta for the same novice-expert participants showed that all three task types demanded increases in arousal and attention porcessing. These same finding were not present in the experienced-expert comparisons. Overall the experienced participants had lower theta present compared to the experts. In fact, experienced participant’s theta levels for the three task types were close to baseline levels, thus indicating that over identically repeated sessions, working memory and concentration task demands are almost nonexistent. This finding needs to be explored further by comparing experts performing these task types at their home plants to see if the drop in theta is a product of repetition of an identical task or familiarity of the environment. Having intimate knowledge of the interface may alleviate some of the task cognitive demands. Beta for experienced participants was no different than experts. This shows that cognitive processing demands for checking detection, and response implementation are similar across the skill continuum when the “equal but different” principle is applied. Specifically, arousal and attention task demands are not alleviated with repeated sessions.

Limitations

This research investigated workload levels and types from novices, experienced, and expert participants when performing a modified EOP using three different task types. The primary goal for this research was to investigate if novices can be used in place of experts using the “equal but different” prinicple. Experienced participants were included in the analysis as a way to capture the skill level continuum; however, this research only had access to two experienced participants so conclusions drawn from the experienced group need to be investigated further to see if the trends hold. Also, the “equal but different” principle attempts to balance novice and experts along the workload dimension. Conclusions drawn outside of workload are likely to be invalid.

Recommendations

This research and its findings have applicability for new NPP control room designs and policy change research. Novice populations are easier to get access to than experts. This research shows that when properly designed, research can use this population as one of the tools for assessing types of workload and to a slightly lesser scope, levels of workload. However, complexity needs to be reduced such that training can account for skill level differences. In addition to novices, experienced participants, over repeated sessions, could provide a ceiling estimate for expert workload levels when operating at their home plant. Another advantage for using novice and experienced participants in workload research is that they are a blank slate for testing new paradigms. Experts are subject to negative transfer effects. This was clearly seen in the different uses of three-way communications by experts.

This research found experts relied on their past training to prioritize tasking such that license critical tasks are completed with more diligence than other less critical tasks. This is seen in the fact that almost all the expert participants identified the correct control on their first attempt; however, several expert participants consistently omitted the recipient with their communications. It was also observed that several participants would request a repeat of an instruction once they found an I&C but before they signaled they had identified the I&C just to double check they had the correct control. Future research should take task priority in to consideration and identify mission critical tasks. Specific training should be provided to novices explaining the importance of getting the task correct even if it means redundant checks.

APPENDIX A: INSTANTANEOUS SELF ASSESMENT RATINGS

Level	Workload	Spare Capacity	Description
1	**Under-utilised**	Very much	Little or nothing to do. Rather boring
2	**Relaxed**	Ample	More time than necessary to complete the tasks. Time passes slowly.
3	**Comfortable**	Some	The controller has enough work to keep him/her stimulated. All tasks are under control.
4	**High**	Very little	Certain non-essential tasks are postponed. Could not work at this level very long. Controller is working 'at the limit'. Time passes quickly.
5	**Excessive**	None	Some tasks are not completed. The controller is overloaded and does not feel in control.

APPENDIX B: NASA TASK LOAD INDEX

NASA Task Load Index

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

APPENDIX C: INFORMED CONCENT NOVICE/EXPERIENCED

Developing a Universal Measure of Workload for Nuclear Power Plant Operators

Informed Consent

Principal Investigator(s): *Lauren Reinerman-Jones, Ph.D.*

Co-Investigator(s):

Sponsor: *Nuclear Regulatory Commission*

Investigational Site(s): *Institute for Simulation and Training*
University of Central Florida
3100 Research Parkway
Orlando, FL 32826

Introduction: Researchers at the University of Central Florida (UCF) study many topics. To do this we need the help of people who agree to take part in a research study. You are being invited to take part in a research study which will include 100 people at UCF. You have been asked to take part in this research study because you are a student at UCF. You must be 18 years of age or older to participate.

The investigator conducting this research is Dr. Lauren Reinerman-Jones from the University of Central Florida's Institute for Simulation and Training.

What you should know about a research study:

- Someone will explain this research study to you.
- A research study is something you volunteer for.
- Whether or not you take part is up to you.
- You should take part in this study only because you want to.
- You can choose not to take part in the research study.
- You can agree to take part now and later change your mind.
- Whatever you decide it will not be held against you.
- Feel free to ask all the questions you want before you decide.

Purpose of the research study: The purpose of this study is to examine first-person experiences while viewing images of space.

What you will be asked to do in the study:

You will be doing task performed my Nuclear Power Plant (NPP) reactor operators on a simulator. When we begin, we will fit you with physiological sensors to monitor your vitals. Throughout and after the experiment you will be taking workload questionnaires.

All of the equipment being used is noninvasive. The devices used in this experiment will be an Advanced Brain Monitoring 10 channel Electroencephalogram (EEG) cap with Electrocardiogram (ECG) sensors attached to it, a BIOPAC 16 channel, 4 light source fNIR strap, and a Spencer Technologies ST3 Digital Transcranial Doppler, with a Marc 600 Headframe.

Each sensor will be custom set for each individual using its respective setup procedure.

The following sections provide a description of the EEG, ECG, fNIR, TCD, and baseline measurement procedure.

EEG: The EEG sensors are contained in a neoprene cap that will be placed over the participant's head and adjusted by the lab technician. The conductive gel is placed on the sensor sponge, which allows the sensor to touch the scalp without being abrasive.

For cap placement, the participant will be seated in front of the computer. The researcher will take an alcohol swab (or equivalent if allergic) and wipe the mastoid bone (behind the ears just above your neck) where the sensors will touch. The research assistant will set the cap so that the front is aligned with the nasium (brow ridge between the eyes) and inion (occipital bone at the back of the head). Once the EEG cap is in place, the research assistant will test the impedance of the sensors to assure that proper conductance is occurring.

ECG: There are two sensors that need to be placed on the right collar bone and the lower left rib bone. These sensors will be placed by the participant. The participant will take an alcohol swab and clean the areas where the sensors will be placed. The research assistant will attach the sensor to the lead and put some conductive gel on the sensor. The participant will then place the sensor in their respective place on the right collar bone or the lower left rib bone. The research assistant will turn on the device and check to see that the EEG and ECG sensors are receiving signal. The signal strength will be evaluated via software on the experimenter's computer station.

fNIR: The fNIR sensors are applied by the research assistant using a strap across the prefrontal cortex. The participant will first wipe their forehead with an alcohol swab and clean the area. Then the fNIR strap will be fitted by the researcher to the participant.

TCD:

The TCD sensors are applied by the research assistant using a head cap that will be placed over the EEG cap that is already on the participants head and adjusted, by the research assistant. The ultrasound gel is placed on the sensors, which allows the sensor to touch the temples without being abrasive.

For cap placement, the participant will be seated in front of the computer. The researcher will take an alcohol swab (or equivalent if allergic) and wipe the participant's temples where the sensors will touch. The research assistant will set the cap on top if the EEG cap. Once the TCD cap is in place, the research assistant will test the signal to assure that proper conductance is occurring.

After the experiment, the research assistant will help you remove all the sensors. It is most beneficial to the research being gathered that you answer all questions and complete all tasks to the best of your abilities, but you are not required to answer every question or complete every task. You will not lose any benefits if you choose not to complete questions or tasks.

Audio or video taping:

You will be audio taped during this study. If you do not want to be audio taped, you will not be able to be in the study. Discuss this with the researcher or a research team member. If you are audio taped, the tape will be kept in a locked, safe place. The tape will be kept indefinitely.

Location: Institute for Simulation and Training, Partnership 2, Room 305.

Time required: We expect that you will be in this research study for 4 hours.

Funding for this study: This research study is being paid for by Nuclear Regulatory Commission.

Risks: There is a small risk that people who take part will develop what is ordinarily referred to as simulator sickness. It occurs once in a while to people who are exposed to prolonged continuous testing in simulated environments. Symptoms consist of nausea and a feeling of being light- headed. The risk is minimized as a result of the short duration of each session in the simulator. If you experience any of the symptoms mentioned, please tell the researcher and remain seated until the symptoms disappear.

All the neurosensing equipment is unobtrusive, non-invasive, and has been fully tested and inspected to maintain safety. The researchers performing this study have completed training on the use and safety of each of the sensors used in the experiment. Because of the conductance gel used in the EEG cap and the ECG sensors, there is a minimal possibility of skin irritation, although the gel is water-based. If this happens, participants are urged to notify the research assistant immediately.

Compensation or payment: Participants may expect to spend 4 hours performing experimental tasks, for which they may elect to receive course credit for the amount of time they participate. Maximum course credit will be 4 credits and is awarded at the discretion of the individual course professor.

Confidentiality: We will limit your personal data collected in this study to people who have a need to review this information. Data will be secured in locked cabinets at the Institute for Simulation and Training (IST) and disposed of following IRB protocol, which includes the shredding of all documents and proper deletion of electronic information.

Study contact for questions about the study or to report a problem: If you have questions, concerns, or complaints, or think the research has hurt you, talk to Dr. Lauren Reinerman-Jones at 407-882-1140 or at lreinerm@ist.ucf.edu.

IRB contact about your rights in the study or to report a complaint: Research at the University of Central Florida involving human participants is carried out under the oversight of the Institutional Review Board (UCF IRB). This research has been reviewed and approved by the IRB. For information about the rights of people who take part in research, please contact: Institutional Review Board, University of Central Florida, Office of Research & Commercialization, 12201 Research Parkway, Suite 501, Orlando, FL 32826-3246 or by telephone at (407) 823-2901. You may also talk to them for any of the following:

- Your questions, concerns, or complaints are not being answered by the research team.
- You cannot reach the research team.
- You want to talk to someone besides the research team.
- You want to get information or provide input about this research.

APPENDIX D: IRB APPROVAL LETTER NOVICE/EXPERIENCED

University of Central Florida

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Lauren Reinerman** and Co-PIs: **Brandon M. Sollins, James L. Tyson IV, Joseph Mercado, Peter A. Hancock, Rebecca Leis**

Date: **June 26, 2013**

Dear Researcher:

On 6/26/2013, the IRB approved the following minor modification to human participant research until 03/20/2014 inclusive:

Type of Review:	IRB Addendum and Modification Request Form
Modification Type:	The time required has changed from 4 to 5 hours and a revised Informed Consent has been approved for use.
Project Title:	Investigating Measures of Workload for Nuclear Power Plant Operators
Investigator:	Lauren Reinerman
IRB Number:	SBE-13-09210
Funding Agency:	NRC
Grant Title:	
Research ID:	64016306

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 03/20/2014, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Signature applied by Joanne Muratori on 06/26/2013 11:04:31 AM EDT

Joanne Muratori

APPENDIX E: INFORMED CONCENT EXPERTS

Examining Measures of Workload for Nuclear Power Plant Operators

Informed Consent

Principal Investigator(s):	*Lauren Reinerman-Jones, Ph.D.*
Co-Investigator(s):	*Daniel Barber, Ph.D.* *Jonathan Harris, M.S.*
Sponsor:	*Nuclear Regulatory Commission*
Investigational Site(s):	*Nuclear Regulatory Commission* *Bethesda, MD*

Introduction: Researchers at the University of Central Florida (UCF) study many topics. To do this we need the help of people who agree to take part in a research study. You are being invited to take part in a research study which will include 70 participants from the Nuclear Regulatory Commission (NRC). You have been asked to take part in this research study because you are a former NPP reactor operator. You must be 18 years of age or older to participate.

The investigator conducting this research is Dr. Lauren Reinerman-Jones from the University of Central Florida's Institute for Simulation and Training.

What you should know about a research study:

- Someone will explain this research study to you.
- A research study is something you volunteer for.
- Whether or not you take part is up to you.
- You should take part in this study only because you want to.
- You can choose not to take part in the research study.
- You can agree to take part now and later change your mind.
- Whatever you decide it will not be held against you.
- Feel free to ask all the questions you want before you decide.

Purpose of the research study: The purpose of this study is to investigate measures of workload in the nuclear power plant domain

UCF University of Central Florida IRB
IRB NUMBER: SBE-16-12319
IRB APPROVAL DATE: 06/07/2016
IRB EXPIRATION DATE: 06/06/2017

What you will be asked to do in the study:

You will be doing tasks performed by Nuclear Power Plant (NPP) reactor operators on a simulator. When we begin, we will fit you with physiological sensors to monitor your vitals. You will be asked to complete workload questionnaires throughout and after the experiment.

All of the equipment being used is noninvasive. The devices used in this experiment will be an Advanced Brain Monitoring 10 channel Electroencephalogram (EEG) cap with Electrocardiogram (ECG) sensors attached to it, a BIOPAC 16 channel, 4 light source fNIR strap, and a Spencer Technologies ST3 Digital Transcranial Doppler, with a Marc 600 Headframe. Each sensor will be custom set for each individual using its respective setup procedure.

The following sections provide a description of the EEG, ECG, fNIR, TCD, and baseline measurement procedure.

EEG: The EEG sensors are contained in a neoprene cap that will be placed over the participant's head and adjusted by the lab technician. The conductive gel is placed on the sensor sponge, which allows the sensor to touch the scalp without being abrasive.

For cap placement, the participant will be seated in front of the computer. The researcher will take an alcohol swab (or equivalent if allergic) and wipe the mastoid bone (behind the ears just above your neck) where the sensors will touch. The research assistant will set the cap so that the front is aligned with the nasium (brow ridge between the eyes) and inion (occipital bone at the back of the head). Once the EEG cap is in place, the research assistant will test the impedance of the sensors to assure that proper conductance is occurring.

ECG: There are two sensors that need to be placed on the right collar bone and the lower left rib bone. These sensors will be placed by the participant. The participant will take an alcohol swab and clean the areas where the sensors will be placed. The research assistant will attach the sensor to the lead and put some conductive gel on the sensor. The participant will then place the sensor in their respective place on the right collar bone or the lower left rib bone. The research assistant will turn on the device and check to see that the EEG and ECG sensors are receiving signal. The signal strength will be evaluated via software on the experimenter's computer station.

fNIR: The fNIR sensors are applied by the research assistant using a strap across the prefrontal cortex. The participant will first wipe their forehead with an alcohol swab and clean the area. Then the fNIR strap will be fitted by the researcher to the participant.

TCD:

The TCD sensors are applied by the research assistant using a head cap that will be placed over the EEG cap that is already on the participants head and adjusted, by the research assistant. The ultrasound gel is placed on the sensors, which allows the sensor to touch the temples without being abrasive.

For cap placement, the participant will be seated in front of the computer. The researcher will take an alcohol swab (or equivalent if allergic) and wipe the participant's temples where the sensors will touch. The research assistant will set the cap on top if the EEG cap. Once the TCD cap is in place, the research assistant will test the signal to assure that proper conductance is occurring.

After the experiment, the research assistant will help you remove all the sensors. It is most beneficial to the research being gathered that you answer all questions and complete all tasks to the best of your abilities, but you are not required to answer every question or complete every task. You will not lose any benefits if you choose not to complete questions or tasks.

Audio or video taping:

You will be audio taped during this study. If you do not want to be audio taped, you will not be able to be in the study. Discuss this with the researcher or a research team member. If you are audio taped, the tape will be kept in a locked, safe place. The tape will be kept indefinitely.

Location: U.S. Nuclear Regulatory Commission, Bethesda, MD.

Time required: We expect that you will be in this research study for 4 hours.

Funding for this study: This research study is being paid for by Nuclear Regulatory Commission.

Risks: There is a small risk that people who take part will develop what is ordinarily referred to as simulator sickness. It occurs once in a while to people who are exposed to prolonged continuous testing in simulated environments. Symptoms consist of nausea and a feeling of being light- headed. The risk is minimized as a result of the short duration of each session in the simulator. If you experience any of the symptoms mentioned, please tell the researcher and remain seated until the symptoms disappear.

All the neurosensing equipment is unobtrusive, non-invasive, and has been fully tested and inspected to maintain safety. The researchers performing this study have completed training on the use and safety of each of the sensors used in the experiment. Because of the conductance gel used in the EEG cap and the ECG sensors, there is a minimal possibility of skin irritation, although the gel is water-based. If this happens, participants are urged to notify the research assistant immediately.

Compensation or payment: Participants may expect to spend 4 hours performing experimental tasks. There is no added compensation, other payment, or extra credit provided to participants for taking part in this study. However, they will be compensated at their hourly rate for their time.

Confidentiality: We will limit your personal data collected in this study to people who have a need to review this information. Data will be secured in locked cabinets at the Institute for Simulation and Training (IST) and disposed of following IRB protocol, which includes the shredding of all documents and proper deletion of electronic information.

Study contact for questions about the study or to report a problem: If you have questions, concerns, or complaints, or think the research has hurt you, talk to Dr. Lauren Reinerman-Jones at 407-882-1140 or at lreinerm@ist.ucf.edu.

IRB contact about your rights in the study or to report a complaint: Research at the University of Central Florida involving human participants is carried out under the oversight of the Institutional Review Board (UCF IRB). This research has been reviewed and approved by the IRB. For information about the rights of people who take part in research, please contact: Institutional Review Board, University of Central Florida, Office of Research & Commercialization, 12201 Research Parkway, Suite 501, Orlando, FL 32826-3246 or by telephone at (407) 823-2901. You may also talk to them for any of the following:

- Your questions, concerns, or complaints are not being answered by the research team.
- You cannot reach the research team.
- You want to talk to someone besides the research team.
- You want to get information or provide input about this research.

APPENDIX F: IRB APPROVAL LETTER EXPERTS

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Lauren Reinerman** and Co-PIs: **Daniel J. Barber, Jonathan T. Harris**

Date: **June 07, 2016**

Dear Researcher:

On 06/07/2016, the IRB approved the following human participant research until 06/06/2017 inclusive:

Type of Review:	UCF Initial Review Submission Form
Project Title:	Examining Measures of Workload for Nuclear Power Plant Operators
Investigator:	Lauren Reinerman
IRB Number:	SBE-16-12319
Funding Agency:	
Grant Title:	Grant number: NRC-HQ-12-C-04-0058
Research ID:	1057461

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 06/06/2017, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Joanne Muratori

Signature applied by Joanne Muratori on 06/07/2016 02:19:02 PM EDT

IRB Manager

REFERENCES

14-Year Cleanup at Three Mile Island Concludes. (1993). Retrieved from http://www.nytimes.com/1993/08/15/us/14-year-cleanup-at-three-mile-island-concludes.html

10075-3, I. (2004). Ergonomic principles related to mental work-load: Part 3: Measurement and assessment of mental work-load: ISO Geneva.

Abich, J. (2013). Investigating the universality and comprehensive ability of measures to assess the state of workload.

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., . . . Craven, P. L. (2007a). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine, 78*(Supplement 1), B231-B244.

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., . . . Craven, P. L. (2007b). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine, 78*(5), B231-B244.

Cole, C. (1993). Shannon revisited: information in terms of uncertainty. *Journal of the American Society for Information Science (1986-1998), 44*(4), 204.

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

Commission, U. S. N. R. (1975). *Reactor safety study. An assessment of accident risks in US commercial nuclear power plants. Executive Summary*. Retrieved from

Commission, U. S. N. R. (2008). *GUIDANCE TO OPERATORS AT THE CONTROLS AND TO SENIOR OPERATORS IN THE CONTROL ROOM OF A NUCLEAR POWER UNIT*. http://www.nrc.gov/reading-rm/adams.html Retrieved from http://www.nrc.gov/docs/ML0823/ML082380236.pdf.

DOE, U. (2009). Human performance improvement handbook: Human performance tools for individuals, work teams and management (Vol. 2; DOE-HDBK-1028-2009). *Washington, DC: Author, Technical Standards Program*.

Dreyfus, S. E., & Dreyfus, H. L. (1980). *A five-stage model of the mental activities involved in directed skill acquisition*. Retrieved from

Estes, S. (2015). The Workload Curve Subjective Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 57*(7), 1174-1187.

Fink, R., Hill, D., & O'Hara, J. (2004). *Human Factors Guidance for Control Room and Digital Human-System Interface Design and Modification. Guidelines for Planning, Specification, Design, Licensing, Implementation, Training, Operation and Maintenance*. Retrieved from

Gawron, V. J. (2008). *Human performance, workload, and situational awareness measures handbook*: CRC Press.

Gevins, A. S., Zeitlin, G. M., Doyle, J. C., Schaffer, R. E., & Callaway, E. (1979). EEG patterns during 'cognitive' tasks. II. Analysis of controlled tasks. *Electroencephalography and Clinical Neurophysiology, 47*(6), 704-710. doi:http://doi.org/10.1016/0013-4694(79)90297-9

Gundel, A., & Wilson, G. F. (1992). Topographical changes in the ongoing EEG related to the difficulty of mental tasks. *Brain Topography, 5*(1), 17-25.

Ha, J. S., Seong, P. H., Lee, M. S., & Hong, J. H. (2007). Development of human performance measures for human factors validation in the advanced MCR of APR-1400. *IEEE Transactions on Nuclear Science, 54*(6), 2687-2700.

Hancock, P., Williams, G., & Manning, C. (1995). Influence of task demand characteristics on workload and performance. *The International Journal of Aviation Psychology, 5*(1), 63-86.

Hancock, P. A., & Meshkati, N. (1988). *Human mental workload*. Amsterdam: North-Holland.

Hankins, T. C., & Wilson, G. F. (1998). A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. *Aviation, space, and environmental medicine, 69*(4), 360-367.

Harris, J., Reinerman-Jones, L., & Teo, G. (2017). The Impact of Simulation Display on Nuclear Power Plant Task Error Frequencies *Advances in Human Factors in Energy: Oil, Gas, Nuclear and Electric Power Industries* (pp. 133-144): Springer.

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

Heiser, S. (2009). Nuclear Industry's Safety, Operating Performance Remained Top-Notch In 2008. Retrieved from Nuclear Street website: http://nuclearstreet.com/nuclear_power_industry_news/b/nuclear_power_news/archive/2009/03/30/nuclear-industry-s-safety-operating-performance-remained-top-notch-in-08-wano-indicators-show#.WApTqCT-maf

Hughes, N., D'Agostino, A., & Reinerman-Jones, L. (2017). The NRC Human Performance Test Facility: An Approach to Data Collection Using Novices and a Simplified Environment *Advances in Human Factors in Energy: Oil, Gas, Nuclear and Electric Power Industries* (pp. 183-192): Springer.

IAEA. (2016, 10-20-2016). Power Ractor Information System. Retrieved from https://www.iaea.org/PRIS/CountryStatistics/CountryDetails.aspx?current=US

Ishihara, S. (2010). *Ishihara's Tests for Colour Deficiency: 24 Plates*: Kanehara Shuppan.

Jarodzka, H., Scheiter, K., Gerjets, P., & Van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction, 20*(2), 146-154.

Joe, J. C., & Boring, R. L. (2017). Using the Human Systems Simulation Laboratory at Idaho National Laboratory for Safety Focused Research *Advances in Human Factors in Energy: Oil, Gas, Nuclear and Electric Power Industries* (pp. 193-201): Springer.

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

John, F. (2008). Three mile island accident. *Docs. school Publications*.

Kim, S., Park, J., Han, S., & Kim, H. (2010). Development of extended speech act coding scheme to observe communication characteristics of human operators of nuclear power plants under abnormal conditions. *Journal of loss prevention in the process industries, 23*(4), 539-548.

Klem, G. H., Lüders, H. O., Jasper, H., & Elger, C. (1999). The ten-twenty electrode system of the International Federation. *Electroencephalogr Clin Neurophysiol, 52*(3).

Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. *Multiple-task performance*, 279-328.

Kubota, Y., Sato, W., Toichi, M., Murai, T., Okada, T., Hayashi, A., & Sengoku, A. (2001). Frontal midline theta rhythm is correlated with cardiac autonomic activities during the performance of an attention demanding meditation procedure. *Cognitive Brain Research, 11*(2), 281-287. doi:http://doi.org/10.1016/S0926-6410(00)00086-0

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.

Lackey, S., Reinerman-Jones, L.E., Salcedo, J. (2014). *Equal but different: 5 research strategies for improving conclusions drawn from novice populations.* Paper presented at the MODSIM World Conference, Hampton, VA.

Larson, A. (2016). A World View of New Nuclear Power Plant Construction. Retrieved from http://www.powermag.com/a-world-view-of-new-nuclear-power-plant-construction-slideshow/

Lee, S. J., Kim, J., & Jang, S.-C. (2011). Human error mode identification for NPP main control room operations using soft controls. *Journal of nuclear science and technology, 48*(6), 902-910.

Leis, R., Reinerman-Jones, L., Mercado, J., Barber, D., & Sollins, B. (2014). *Workload from Nuclear Power Plant Task Types Across Repeated Sessions.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Lin, C. J., Yenn, T.-C., & Yang, C.-W. (2010). Automation design in advanced control rooms of the modernized nuclear power plants. *Safety science, 48*(1), 63-71.

Luque-Casado, A., Zabala, M., Morales, E., Mateo-March, M., & Sanabria, D. (2013). Cognitive Performance and Heart Rate Variability: The Influence of Fitness Level. *PLoS ONE, 8*(2), e56935. doi:10.1371/journal.pone.0056935

Matthews, G., Reinerman-Jones, L., Wohleber, R., Lin, J., Mercado, J., & Abich IV, J. (2015). *Workload Is Multidimensional, Not Unitary: What Now?* Paper presented at the International Conference on Augmented Cognition.

Matthews, G., Reinerman-Jones, L., Wohleber, R., Lin, J., Mercado, J., & Abich, J. (2015). Workload Is Multidimensional, Not Unitary: What Now? In D. D. Schmorrow & C. M. Fidopiastis (Eds.), *Foundations of Augmented Cognition: 9th International Conference, AC 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings* (pp. 44-55). Cham: Springer International Publishing.

Mercado, J. (2014). Assessing the effectiveness of workload measures in the nuclear domain.

Mercado, J. E., Reinerman-Jones, L., Barber, D., & Leis, R. (2014). Investigating Workload Measures in the Nuclear Domain. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58*(1), 205-209. doi:doi:10.1177/1541931214581043

Mercado, J. E., Reinerman-Jones, L., Barber, D., & Leis, R. (2014). *Investigating Workload Measures in the Nuclear Domain.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review, 63*(2), 81.

Min, D., Chung, Y. H., & Yoon, W. C. (2004). Comparative analysis of communication at main control rooms of nuclear power plants.

Miyake, S. (2001). Multivariate workload evaluation combining physiological and subjective measures. *International journal of psychophysiology, 40*(3), 233-238.

Moray, N. (2013). *Mental workload: Its theory and measurement* (Vol. 8): Springer Science & Business Media.

Mulder, L., de Waard, D., & Brookhuis, K. A. (2004). Handbook of human factors and ergonomics methods.

Mulder, L. J., de Waard, D., & Brookhuis, K. A. (2004). Estimating mental effort using heart rate and heart rate variability *Handbook of human factors and ergonomics methods* (pp. 20-21-20-28): CRC Press.

Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 510.

NRCgov (Writer). (2013). Moments in NRC History: Three Mile Island - March 28,1979.

O'Hara, J., Brown, W., Lewis, P., & Persensky, J. (2002). Human-system interface design review guidelines (NUREG-0700, Rev 2). *Washington, DC: US Nuclear Regulatory Commission*, 12.

O'Hara, J., & Higgins, J. (2010). Human-system interfaces to automatic systems: Review guidance and technical basis. *Human Factors of advanced reactors (NRC JCN Y-6529) BNL Tech Report No BNL91017-2010*, 1-23.

Park, J., Jung, W., & Yang, J.-E. (2012). Investigating the effect of communication characteristics on crew performance under the simulated emergency condition of nuclear power plants. *Reliability Engineering & System Safety*, *101*, 1-13.

Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., & Pope, A. T. (2000). A Closed-Loop System for Examining Psychophysiological Measures for Adaptive Task Allocation. *The International Journal of Aviation Psychology, 10*(4), 393-410. doi:10.1207/S15327108IJAP1004_6

Raeisi, S., Osqueizadeh, R., Maghsoudipour, M., & Jafarpisheh, A. (2016). Ergonomic redesign of an industrial control panel. *The International Journal of Occupational and Environmental Medicine, 7*(3 July), 756-186-792.

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-13*(3), 257-266. doi:10.1109/TSMC.1983.6313160

Reinerman-Jones, L., Guznov, S., Mercado, J., & D'Agostino, A. (2013). *Developing methodology for experimentation using a nuclear power plant simulator.* Paper presented at the International Conference on Augmented Cognition.

Reinerman-Jones, L., Hughes, N., & D'Agostino, A. (2016). *A Comparison of Displays and Associated Workload on Nuclear Power Plants Tasks.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Reinerman-Jones, L., Mercado, J., D'Agostino, A., Hughes, N., Sollins, B., & Leis, R. (2015). Workload Associated with Nuclear Power Plant Main Control Room Tasks. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 59*(1), 110-114. doi:doi:10.1177/1541931215591023

Reinerman, L. E., Matthews, G., Warm, J. S., Langheim, L. K., Parsons, K., Proctor, C. A., . . . Stutz, R. M. (2006). Cerebral Blood Flow Velocity and Task Engagement as Predictors of Vigilance Performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*(12), 1254-1258.

Seong, P. H., Kang, H. G., Na, M. G., Kim, J. H., Heo, G., & Jung, Y. (2013). Advanced MMIS toward substantial reduction in human errors in NPPs. *Nuclear Engineering and Technology, 45*(2), 125-140.

Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review, 5*(1), 3-55.

Simon, H. A. (1962). The architecture of complexity *Facets of Systems Science* (Vol. 106, pp. 457-476): Springer.

So, H., & Chan, K. (1997). *Development of QRS detection method for real-time ambulatory cardiac monitor.* Paper presented at the Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE.

Systems, G. (2015). GSE'S GENERIC NUCLEAR PLANT SIMULATOR (pp. 1-4): GSE Systems.

Systems, G. P. (2011). GPWR Nuclear Plant Operating Manual *Loss of AC Power to 1A – SA and 1B – SB Buses* (Vol. 3 part 4).

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

Tanner, W. P., Jr., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological review, 61*(6), 401-409. doi:10.1037/h0058700

Thakor, N. V., & Tong, S. (2004). Advances in quantitative electroencephalogram analysis methods. *Annual Review of Biomedical Engineering, 6*, 453-495.

Three Mile Island Accident. (2013). Retrieved from http://www.nrc.gov/reading-rm/doc-collections/fact-sheets/3mile-isle.pdf

Training Applications | GSE Systems. (2015). Retrieved from http://www.gses.com/training-applications

Turner, J. R. (1994). *Cardiovascular reactivity and stress: Patterns of physiological response*. New York, NY, US: Plenum Press.

Vygotsky, L. (1987). Zone of proximal development. *Mind in society: The development of higher psychological processes, 5291*, 157.

Wertheim, A. H. (1981). Occipital Alpha Activity as a Measure of Retinal Involvement in Oculomotor Control. *Psychophysiology, 18*(4), 432-439. doi:10.1111/j.1469-8986.1981.tb02476.x

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2015). *Engineering psychology & human performance*: Psychology Press.

Williges, R. C., & Wierwille, W. W. (1979). Behavioral Measures of Aircrew Mental Workload. *Human Factors, 21*(5), 549-574.

Wilson, G. F., & Eggemeier, F. T. (1991). Psychophysiological assessment of workload in multi-task environments. In D. L. Damos (Ed.), *Multiple-task performance* (pp. 329–360). Washington, DC: Taylor & Francis.

Yeh, Y.-Y., & Wickens, C. D. (1988). Dissociation of Performance and Subjective Measures of Workload. *Human Factors*, *30*(1), 111-120.