Assessing the Effectiveness of Workload Measures in the Nuclear Domain

Joseph Mercado
University of Central Florida

ASSESSING THE EFFECTIVENESS OF WORKLOAD MEASURES IN THE NUCLEAR DOMAIN

by

JOSEPH E. MERCADO

B.A. Mercy College, 2007
M.S. Mercy College, 2010

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Modeling and Simulation – Human Systems
in the Institute for Simulation and Training
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Summer Term
2014

Major Professor: Lauren Reinerman-Jones and Peter Hancock

ABSTRACT

An operator's performance and mental workload when interacting with a complex system, such as the main control room (MCR) of a nuclear power plant (NPP), are major concerns when seeking to accomplish safe and successful operations. The impact of performance on operator workload is one of the most widely researched areas in human factors science with over five hundred workload articles published since the 1960s (Brannick, Salas, & Prince, 1997; Meshkati & Hancock, 2011). Researchers have used specific workload measures across domains to assess the effects of taskload. However, research has not sufficiently assessed the psychometric properties, such as reliability, validity, and sensitivity, which delineates and limits the roles of these measures in workload assessment (Nygren, 1991). As a result, there is no sufficiently effective measure for indicating changes in workload for distinct tasks across multiple domains (Abich, 2013). Abich (2013) was the most recent to systematically test the subjective and objective workload measures for determining the universality and sensitivity of each alone or in combination. This systematic approach assessed taskload changes within three tasks in the context of a military intelligence, surveillance, and reconnaissance (ISR) missions. The purpose for the present experiment was to determine if certain workload measures are sufficiently effective across domains by taking the findings from one domain (military) and testing whether those results hold true in a different domain, that of nuclear. Results showed that only two measures (NASA-TLX frustration and fNIR) were sufficiently effective at indicating workload changes between the three task types in the nuclear domain, but many measures were statistically significant. The results of this research effort combined with the results from Abich (2013) highlight an alarming problem. The ability of subjective and physiological measures to indicate

changes in workload varies across tasks (Abich, 2013) and across domain. A single measure is not able to measure the complex construct of workload across different tasks within the same domain or across domains. This research effort highlights the importance of proper methodology. As researchers, we have to identify the appropriate workload measure for all tasks regardless of the domain by investigating the effectiveness of each measure. The findings of the present study suggest that responsible science include evaluating workload measures before use, not relying on prior research or theory. In other words, results indicate that it is only acceptable to use a measure based on prior findings if research has tested that measure on the exact task and manipulations within that specific domain.

This work is dedicated to my family. Especially to my wife Bia and son Jordan, you are my motivation.

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.

I would like to thank the members of my committee: Dr. Stephanie Lackey, Dr. James Szalma, and especially Dr. Lauren Reinerman-Jones and Dr. Peter Hancock, for serving as my academic advisors throughout my graduate career. I have grown tremendously under your tutelage and I appreciate all of your guidance and encouragement. I would also like to that my Army Research Institute mentor Dr. Randall Spain, you truly helped me grow as a researcher.
I would like to thank the IST's ACTIVE lab, with a special thanks to the members that helped contribute to this work, specifically Dr. Daniel Barber, Sgt. James Tyson, Ryan Wohleber, Sushi Martinez, Sherry Ogreten, Brandon Sollins, Jinchao Lin, and Rebecca Leis. Lastly, I would like to thank my family for all of the encouragement and support throughout the years. To my wife Bia, thank you for being there every step of the way, you pushed me to become a better student when we first met and this is your doctorate degree as much as it is mine. To my son Jordan, thank you for giving me to motivation to work hard every day and for being the one person who took all of my stress away. To my grandparents, thank you for making my goals and dreams your goals and dreams. To my father, thank you for push me to succeed and constantly telling me to strive for greatness.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

An operator's performance and mental workload when interacting with a complex system, such as the main control room (MCR) of a nuclear power plant (NPP), are major concerns in seeking to accomplish safe and successful operations. The impact of performance on operator workload is one of the most widely researched areas in human factors science, with over five hundred workload articles published since the 1960s (Brannick, Salas, & Prince, 1997; Meshkati & Hancock, 2011). Human operators working in domains such as medicine, aviation, military, and nuclear technologies, face many challenges when performing critical tasks requiring complex systems that impose varying levels of demand (Abich, 2013; Huey & Wickens, 1993; Stanton, Salmon, Walker, Baber, & Jenkins, 2010). A complex system is composed of electrical parts, chemical parts, mechanical parts, and/or a combination thereof that interconnect to lead to functionality for a task (Brown, Conrad, & Beyeler, 2012). Interaction with multiple complex systems sometimes overwhelms the operator due to its high rate of information flow (Paas & Merriënboer, 1994). Organizations have become aware of this problem and more specifically, the cost to business and safety of this problem. Thus, they seek to lower the level of demand placed on the operator (Hwang et al., 2008). To effectively lower demand while maintaining performance, measures of workload are necessary to understand the relation between performance and taskload. System design and task requirements are only as good as the metrics that determine their development.

Researchers have used specific workload measures, such as the National Aeronautics and Space Administration-Task Load Index (TLX), Instantaneous Self-Assessment (ISA), and Multiple Resource Questionnaire (MRQ), across domains to assess the effects of taskload.

However, research has not sufficiently assessed the psychometric properties, such as reliability, validity, and sensitivity, which delineates and limits the roles of these measures in workload assessment (Nygren, 1991). Many of these measures were intended for a specific domain or for only a specific task. Frequently, this results in research using measures in domains for which they were never validated in. For example, Hart and Staveland (1988) developed the NASA-TLX the aviation, but it has become the most commonly used workload measure, crossing multiple domains. There is no universal sufficiently effective measure for indicating changes in workload for distinct tasks across multiple domains (Abich, 2013). Abich (2013) was the most recent to systematically test the three aforementioned workload measures and others for determining the universality and sensitivity of each alone or in combination. This systematic approach assessed taskload changes within three tasks in the context of a military intelligence, surveillance, and reconnaissance (ISR) mission. The purpose for the present experiment is to determine if certain workload measures are sufficiently effective across domains by taking the findings from one domain (military) and testing whether those results hold true in a different domain, that of nuclear.

Workload Defined

Workload is a result of taskload and performance on a task. The initial workload theory was unitary resource theory (Kahneman, 1973; Moray, 1967) and subsequently multiple resource theory (MRT; Wickens, 1984, 1992, 2008). Both of these postulate the idea that the human system possesses a finite amount of cognitive resources. The major difference between each theory lies in the constraints on such resources (Kantowitz & Knight, 1976; Navon & Gopher, 1979; Wickens, 1976). Resource theory argues that humans possess one central pool of resources, while MRT argues humans possess different resource pools with varying capacities (Kahneman, 1973; Moray, 1967; Wickens, 1980, 1984; Wickens & Hollands, 2000).

In terms of resource theory, workload occurs as a result of the amount of resources allocated to a task and when such resources are depleted. Workload increases and performance suffers. Cognitive resources drawn from this unitary pool might include mental, physical, effort, frustration, and more. Regardless of the type of cognitive resource, the unitary pool is used to meet the demand of the tasks at hand. Therefore, workload will increase and performance will suffer even if tasks are drawing on different types of resource (Friedenberg & Silverman, 2006; Kahneman, 1973). MRT asserts that workload is not increased until one or more pools of resources are depleted and this does not necessarily mean that all of workload increases. For example, the depletion of the verbal processing pool might not be accompanied by the utilization of the spatial pool of resources. Therefore, if resource demand is equal, two tasks that both demand one level of a given dimension will inhibit each other more than two tasks that require separate levels on the same dimension (Wickens, 2002).

Workload can thus be explained in one of two approaches. Thus, a single definition of workload is not agreed upon. However, all proposed definitions stem from two fundamental themes. First, all proposed definitions consider workload as an active interaction between the operator and their task (Megaw, 2005). Second, all proposed definitions theorize workload as the amount of information processing, mental effort, and/or cognitive resources required for task performance, relative to their capacity (Abich, 2013; Eggemeier, Wilson, Kramer, & Damos, 1991; Gopher & Donchin, 1986; Hockey, 1997; Kahneman, 1973; Kramer, Sirevaag, & Braune, 1987; Moray, 1979; Taylor, 2012; Veltman & Gaillard, 1996). Various definitions of workload exist (Gopher & Donchin, 1986; Navon & Gopher, 1979), but the working definition examined in the present paper is “workload might be the operators 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).

Assessing Workload

Within the last four decades, the applied community has expressed substantial and continuing interest in the concept of workload. The main reason for measuring workload is to assess the mental cost of performing tasks and predict operator and system performance (Cain, 2007). Research on workload has sought to answer questions such as: “How busy is the operator?” and “Will the operator be able to respond to an unexpected event?” (Wickens & Hollands, 2000b). No single general measure of workload exist (Gopher & Donchin, 1986). Thus, an operator’s workload when interacting with a complex system, such as the MCR of a NPP, has been assessed using a number of measures. The present experiment assesses operator workload via the three broad categories identified by Eggemeier et al. (1991): subjective rating scales (self-assessment), and two objective forms of measure (performance and physiological)[1].

Task demands might be multi-dimensional, yet it is unknown whether an operator’s conscious perception of workload is best characterized by a multidimensional approach or by a scalar measure (Cain 2007). As a result, subjective measures are typically in the form of questionnaires that are founded on either resource theory (Kahneman, 1973; Moray, 1967;) or MRT (Wickens, 1984, 1992, 2008). Performance measures come in the form of primary and secondary task performance, where decrements indicate a change in workload (Wickens &

[1] Reference Abich (2013) and Cain (2007) for a detailed explanation of subjective and objective measures of workload

Hollands, 2000b). Physiological measures continuously monitor bodily responses to associated changes in taskload (Cain, 2007).

Subjective Measures of Workload.

Subjective measures are highly applicable to assessing an operator's workload when interacting with modern technologies that aid judgment and decision making (Cain, 2007), such as technologies in the military and nuclear domains. These measures evaluate an operator's interpretations and judgments of their experienced task demand. Subjective measures are the most commonly used method to measure an operator's workload because they are easy to administer, analyze, and complete. Subjective measures are nonintrusive to primary task performance because they are typically collected post-task (Wickens & Hollands, 2000b). When compared to objective measures of workload, some argue that an operator's perceived workload demand (subjective) is a more effective measure of workload because it is sensitive to minor changes in task demand (Johannsen, 1979; Muckler & Seven, 1992). Thus, there appears to be very few costs associated with subjective measures.

However, subjective measures have limitations. These measures depend on an operator's perception and are typically collected post-task. Measures that rely on an operator's perception are subject to operator bias and are only based on what an operator remembers from their experience (Cain, 2007). Post-task measures of workload typically cause operator's to forget where workload changes occurred during the task. The alternative of applying these measures during the task eliminates this disconnect, but then the measure can interfere with task performance (Hockey & Tattersall, 1995).

The present experiment assesses subjective workload via the same three measures (TLX, ISA, and MRQ) used by Abich (2013), which have been applied extensively across many domains and through various methods of administration. The three subjective workload measures cover both post-task and in-task (online) collection options.

NASA- TLX (Task Load Index).

Founded on resource theory, the TLX is the most commonly used subjective workload rating scale. Administered post-task, the TLX evaluates global workload and six subscales of workload that includes mental demand, physical demand, temporal demand, performance, effort, and frustration (Hart & Staveland, 1988).

ISA.

Ascribing to resource theory, the ISA is a subjective uni-dimensional workload rating method. The ISA provides an online assessment of task demand on perceived workload (Tattersall & Foord, 1996). In most instances, this rating is collected during the task via an auditory prompt.

MRQ.

Founded on MRT, the MRQ is used to characterize the nature of the mental processes used during a task (Boles & Adair, 2001). This post-task questionnaire includes seventeen scales that are based on factor analytic studies of lateralized processes (Boles, 1991, 1992, 1996, 2002), which suggest that subjective mental workload measures based on resource theory (Kahneman, 1973; Moray, 1967;) need to expand to assess a wider range of mental processes (Boles et al., 2001).

Objective Measures of Workload.

Performance.

Both resource theory and MRT state that an operator has a limited capacity of resources to allocate towards the demands of a task (Kahneman, 1973; Moray, 1967; Wickens, 2008; Wickens & Hollands, 2000). Therefore, measuring task performance should provide an indication of workload, in theory. This approach is particularly useful when the task demands exceed the operator's capacity such that performance degrades (Eggemeier, et al., 1991). The present experiment measures primary task performance, which assesses the operator's performance on the task of interest directly.

Primary task performance measures are nonintrusive, since they are the focus of participants. Measuring primary task performance is typically easy as it allows for continuous data collection. Commonly used measures of primary task performance include, speed, accuracy, response time, and error rate (Cain, 2007; Hancock, Mercado, Merlo, Van Erp, 2013; Mercado, White, Sanders, Wright, Hancock, 2013; O'Donnell & Eggemeier, 1986; Paas & Van Merriënboer, 1993). However, criteria should be dependent upon the task domain. For example, error rate, as opposed to correct responses, could provide more insight to operational tasks (Cain, 2007).

Using primary performance data as a measure of workload has its limitations. Factors including skill, experience, practice effects, and training often affect the efficiency of performance at high task demand (Hinds, 1999). Low task demand is affected by boredom

(Hart, 2010). Hart and Wickens (1990) state that primary task performance is more a measure of what the system can achieve and not an accurate estimate of the cost of operator achievement. As a result, dissociation between workload and primary task performance is frequently observed (Yeh and Wickens, 1988) and some performance measures alone cannot describe workload. Furthermore, primary task performance measures are difficult to standardize across domains because different performance measures are required from different domains (Meshkati and Lowewinthal, 1988). These limitations lead performance measures to typically be combined with subjective and physiological measures (Miller, 2011).

Physiological Measures of Workload.

Wierwille (1988) argues that subjective measures alone are inadequate to effectively characterize workload because they can become insensitive to changes in task demand. He suggests additional measures for capturing instantaneous or real time workload are necessary. One way to achieve this is through physiological measures. The primary appeal of physiological measures is their continual and objective measurement of an operator’s state. Past research suggests physiological measures correlate well with various aspects of workload, and are seen as promising objective workload measures (Cain, 2007).

Unlike subjective questionnaires that capture perceptual responses to taskload, physiological measures of workload record physiological responses to changes in task demand (Hess & Polt, 1964; O’Donnell & Eggemeier, 1986; Rasmussen, 1979). The goal when using these measures is to develop assessments with well-known properties that can be applied in specific situations (Cain, 2007).

Using physiological sensors for measuring of workload has its limitations. It is unclear as to whether or not physiological measures assess workload as imposed by a task (Meshkati et al., 1995). Instead, they provide information regarding how the operator responds and copes to taskload. In addition, there is a degree of invasiveness with some physiological measures but advances in technology have considerably reduced this burden (Cain, 2007).

Wilson and O'Donnell (1988) note that one particular physiological measure is unlikely to universally measure workload because of the complex nature that the construct of workload has evolved to embody. Currently, physiological experts suggest that a battery of physiological measures be used when investigating mental workload (Cain, 2007). As a result, the present experiment uses several physiological measures including Electroencephalography (EEG), Transcranial Doppler (TCD) ultra-sonography, functional Near Infra-Red (fNIR), and Electrocardiography (ECG).

EEG.

EEG is a direct measure of the neural activity by recording electrical activity of the brain with electrodes placed on the scalp of the operator. EEG is sensitive to changes in mental workload and the cognitive tasks performed (Brookings, Wilson & Swain, 1996; Taylor, Reinerman-Jones, & Cosenzo, & Nicholson, 2010).

TCD.

TCD monitors cerebral blood flow velocity (CBFV) in intracranial arteries (Tripp & Warm, 2007). Ultra-sonography technology, similar to the ultra-sound technology used in prenatal care, is used to capture the CBFV in all arteries of both left and right hemispheres.

Mental workload is frequently measured by increased CBFV in regions of the prefrontal cortex, specifically the medial cerebral arteries (Parasuraman & Caggiano, 2005; Reinerman-Jones, Matthews, Langheim, & Warm, 2011).

fNIR.

fNIR is used to monitor (hemodynamic) changes in oxygenated hemoglobin (oxy-Hb) and deoxygenated hemoglobin (deoxy-HB) in the brain, i.e., cerebral hemodynamic response (Ayaz et al., 2010; Ayaz et al., 2011; Chance, Zhuange, UnAh, Alter, & Lipton, 1993;). In the present experiment, the fNIR will measure oxygenation in the prefrontal cortex. Increases in blood oxygenation have been shown when task difficulty increases (Ayaz et al., 2010).

ECG.

ECG is a direct measure of cardiac activity and one of the most frequently used physiological measures of workload. Decreases in Inter-beat Interval (IBI) have been linked with increased mental workload (Veltman & Gaillard, 1996). Heart Rate Variability (HRV) reflects engagement in effortful information processing. Increases in workload have also been linked with increases in heart rate (Jorna, 1993; Veltman & Gaillard, 1996; Wilson, Fullenkamp, & Davis, 1994).

Justification

Regardless of the battery of workload measures selected, formulation of a general measure of workload requires multiple varied experiments; results from a single experiment are insufficient (Wierwille, 1988). Thus, the extensive research on measuring workload has yet to lead to a generalizable measure of workload because the focus has been on understanding taskload for maintaining or improving performance, not on the actual measurement tools. The purpose for the present experiment is to determine the generalizability of certain workload measures by taking the findings from one domain (military) and testing whether those results hold true in a different domain, that of nuclear. In order to compare workload measures across domains, it is first important to review previous workload research in the military and nuclear domains and to explain the composition of tasks required for successful operation within these domains.

Workload in the Military Domain (ISR).

ISR assists decision making of military commanders through the incorporation and synchronization of battlefield operating systems to gather and process intelligence about the enemy and activities. Presently, both military commanders and policymakers in Washington D.C. use ISR systems to track developments in combat zones (Erwin, 2013). In an ISR mission, operators perform tasks that require the detection of threats and changes within the operational environment. More often, operators perform multiple tasks at the same time (combined threat and change detection). Multitasking occurs when a person performs two or more tasks simultaneously or in rapid succession (Gopher, Armory, & Greenspan, 2000). Multitasking stems from the idea that humans have a “execute control” that has two separate balancing stages. The “goal shifting” stage allows the person to choose the task they want to do and the “rule activation” stage allows the person to turn off the rules for one task and turn on the rules for the other task (Meyer et al., 1997).

Threat and change detection tasks occur within the same domain, yet they are theoretically different. Threat detection tasks stem from signal detection theory (SDT), which theorizes that most decision making occurs in the face of ambiguity because noise is continuously present (Green & Swets, 1996; Heeger, 1997). Noise can be both internal (perceptual processing and neural activity) and external (environmental factors). An operator's performance is determined by how well they can discriminate the signal from the noise (Wickens & Holland, 2000). In threat detection tasks, there are four possible results: hit, miss, false alarm, or correct rejection. A hit occurs when an operator decides a signal exists and it does. A miss occurs when a signal exists, but the operator does not notice. A false alarm occurs when an operator decides a signal exists, but it does not. A correct rejection occurs when there is no signal and an operator decides there is no signal.

Change detection tasks stem from change detection and change blindness theories. Change detection is the visual process that occurs when noticing a change that calls for detection, identification, and localization of a stimulus. This process answers the questions: did a change occur, what kind of change was it, and where did the change occur (Rensick, 2002). Change blindness occurs when an observer fails to notice a change in a visual scene (Rensink, 1997; Simon, 1996; Simon & Ambinder, 2005; Simon & Levin, 1997).

Abich (2013) was the first study to assess systematically the sensitivity and comprehensiveness of subjective, performance, and physiological workload measures used together within the military (ISR) domain for theoretically different tasks. This study focused on the three tasks that are vital in ISR military operations: threat detection, change detection, and multitasking (combined threat and change detection). The threat detection task demand comprised of three conditions that manipulated event rate: low (15), medium (30), and high (60) events per minute. These three event rates were coupled with a medium threat probability of (2:15), resulting in three ratios: 2:15, 4:30, and 8:60. The change detection task demand was also comprised of three conditions: low (6), medium (12), and high (24) events per minute. These three event rates were coupled with a medium signal saliency that consisted of two icons changing simultaneously, resulting in three ratios: 2:6, 2:12, and 2:24. Event rates for both tasks were resultant from a pilot experiment (Abich, 2013), but were originally derived from previous work (See, Howe, Warm, & Dember, 1995; Taylor, 2012). One ISR mission consisted of four scenarios: change detection task, threat detection task, change detection task with threat detection task held at a constant level (medium event rate), and threat detection task with change detection held at a constant level (medium event rate). The subjective workload measures used in this experiment were TLX, ISA, and MRQ. The physiological workload measures used were EEG, TCD, fNIR, ECG, and eye tracking. Results suggest that both subjective and objective measures are sensitive to differences in workload associated with task demand when examining the effects of event rate on both a signal and change detection task within a complex military operation. However, sensitivity varied within and across measures.

In regards to subjective workload, Abich (2013) highly recommends the ISA for a global, on-line assessment because it shows negligible interference with task performance. ISA was sensitive to task demand for every task. When comparing both post-task subjective measures (TLX and MRQ), the MRQ was superior because it detected dimensions affecting workload that the TLX could not identify and was selective in measuring the effects of task demand on performance. All TLX subscales were sensitive to task demand for all tasks, but the TLX only has six subscales. Whereas the MRQ showed that seven out of the ten scales used were sensitive to task demand for the threat detection, change detection, and threat detection with change detection held constant tasks. However, only three out of the ten scales used were sensitive to task demand for the change detection with threat detection held constant task.

The physiological results shown by Abich (2013) were not as clear as the subjective results. However, the experiment revealed interesting findings for all physiological measures. For the purposes of generalizing these results for testing in the present experiment, eye tracking is not discussed because that sensor was not included due to technical limitations. Of the remaining sensors, ECG was found to yield the most promising measure of workload because it was sensitive to task demand for all tasks. Abich (2013) recommends ECG, specifically HRV, because of his results combined with the ECG's economical cost and straightforward application. EEG yielded mixed results, as Alpha F4 and beta Fz were the only measures sensitive to task demand for all tasks. TCD and fNIR results were less promising, as they were only sensitive to changes during one task, which does not provide convincing support for their use as a workload measure of the task demand levels implemented by Abich (2013).

Workload in the Nuclear Domain (NPP MCR Operations).

The purpose for an NPP is to generate electricity from steam created by nuclear heat. The electricity produced by the 100 commercial NPP reactors in the United States is equivalent to 31% of the world's total nuclear-generated electrical power (U.S. NRC, 2013). There are two types of NPP's used in the United States: boiling water reactor (BWR) and pressurized water reactor (PWR). In BWRs, which account for one-third of the commercial power reactors in the United States, the reactor core heats water that turns to steam, which powers a steam turbine. In PWRs, which account for two-thirds of the commercial power reactors in the U.S., the reactor core heats water, but not to boiling point. This hot water then exchanges heat with a lower pressure water system, which turns to steam that powers the turbine (U.S. NRC, 2013). Even though NPPs provide cost effective electricity, they bring concerns related to the health impact and safety of the public, specifically exposure to pollution by-products. The effects of radiation on a human can be terrifying and mysterious. As a result, safe operation is of the utmost important. The key personnel in NPP operations are ROs. Their responsibility is to supervise the NPP and perform actions to safeguard the NPP.

A NPP is a complex system controlled through a Human System Interface (HSI) located in the MCR (Reinerman-Jones, Guznov, Mercado, & D'Agostino, 2013). Two types of reactor operators manage and maintain a NPP MCR, RO's and Senior Reactor Operator (SRO). In a highly automated NPP, the most common tasks performed by operators are monitoring instrumental panels and detecting the state of the NPP. Monitoring and detection is one of the four primary tasks performed by both ROs and SROs outlined by O'Hara and colleagues (2008, 2010). Monitoring requires checking the plant to determine whether it is functioning properly by verifying parameters indicated on the control panels, observing the readings displayed on screens, and obtaining verbal reports from other personnel. Detection occurs when the operator recognizes the state of the plant has changed. The three other primary tasks are situational assessment, response planning, and response implementation. Situational assessment tasks consist of evaluating current states of NPP systems to ensure they are within required parameters or to determine the underlying cause of any irregularities. Response planning tasks consist of deciding on a plan of action to diagnose and perform appropriate actions at the NPP and are guided by standardized symptom-based procedures called Emergency Operating Procedures (EOPs). Response implementation tasks consist of performing actions required by response planning (e.g., selecting a control, performing action on the control, and monitoring responses of the system and process (O’Hara et. al., 2008; O'Hara & Higgins, 2010). O’Hara and colleagues provide a great starting point for identifying the tasks performed by reactor operators, but additional refinement is necessary. A tasks analysis revealed the need for an additional task identified as checking and the redefinition of the other tasks (Reinerman-Jones et al., 2013). The checking task requires a one-time inspection of an instrument or control to verify that it is in the appropriate state. At its foundation, the checking task is a successive-attention task, requiring participants to retain critical information in their working memory and distinguish an indicator from a non-indicator (Reinerman, 2006). The detection task requires continuous monitoring of a control parameter for identification of change. This task stems from SDT, requiring participants to remain vigilant to discriminate a signal from noise. The response implementation task requires an action to affect the state of the NPP. Response implementation is a fine motor response task, requiring participants to use a mouse and turn a switch.

In conjunction with performing, the tasks detailed above ROs must maintain proper three-way communication, as a way of relaying critical information. Three-way communication is a method for relaying information and checking for understanding between team members by clearly and simply expressing all components of the communication and confirming instructions. Three-way communication is how the SRO communicates task instructions to each RO. For each tasks instruction, three-way communication contains two three-way parts. Both parts require the same three pieces of instruction. First is the initiation of the instruction. Second is the understanding of the initiated instruction. Third is the confirmation of the comprehension statement.

Advancements, such as analog to digital, in human-system interfaces (HSIs) in NPP MCRs have changed the role of ROs, which can result in performance and safety concerns.

These advancements combined with the heightened awareness that NPP incidents occur because of an interaction between the human operator and the complex system have caused a growing need to incorporate human factors principles in NPP operations (Reinerman-Jones, Guznov, Tyson, & D'Agostino, 2012). As a result, the research into human-complex system interaction has intensified in the nuclear domain (Lin et al., 2011).

To uphold safe and efficient NPP operation, it is vital to measure an operator's workload during system operation (Hwang et al., 2008). Past research on workload in the NPP domain has focused on modifying existing metrics, such as the TLX, by altering instructions and the amount of items included. Other research has attempted to use physiological measures, but have experimentation flaws due to poor experimental design and implementation.

Lin et al. (2011) compared the effectiveness of the NASA-TLX and Team Workload Assessment (TWA) in measuring team workload during an EOP. The TWA, which is modified from the NASA-TLX, considers teamwork to be a four-element process including coordination, communication, leadership and support, and time-sharing. The findings of Lin et al. (2011) suggest that the TWA is more sensitive to task performance when compared to the NASA-TLX for assessing NPP RO crews.

Other research in the NPP domain has focused on the effects of automation on mental workload. Lin, Yenn, and Yang (2010) examined the effects of different levels of automation (LOAs) under different operating procedures on operator performance in a NPP MCR. The two operating procedures used were an integrated operating procedure (IOP) and an abnormal operating procedure (AOP). The LOAs used for the IOP were action support (LOA 2) and supervisory control (LOA 9). The action support LOA required the operator to generate a target processing order. The supervisory support LOA presented a computer generated processing strategy that the operator could override. The LOAs used for the AOP were shared control (LOA 5) and blended decision-making (LOA 6). In the shared control LOA, the computer generated the decisions that the operator could select or edit. In the blended decision-making LOA, the computer generated decision options, but required the operator to select the one they felt was the best. The experimental task used was a modified version of the Personal Computer Transient Analyzer (PCTRAN) system and alarming processing system used by Huang et al. (2006). The results showed a significant difference in mental workload between different levels of (LOAs), with the blended-decision-making eliciting the lowest mental demand.

Hwang et al. (2008) examined mental workload and performance of diagnosis and monitoring tasks in the MCR of an NPP. The experiment consisted of a simulated reactor shutdown task with a secondary task that required the operator to assess the relation between performance, mental workload, and physiological measures. The research goal was to develop a closed-loops system using group method data handling, seven physiological measures, and secondary task performance. The seven physiological measures used included: parasympathetic/sympathetic ratio, heart rate, heart rate variability, diastolic pressure, systolic pressure, eye blink frequency, and eye blink duration. The primary task consisted of shutting down the reactor and the secondary task consisted of mental arithmetic problems. Results show a positive correlation between NASA-TLX and error rate. In addition, all physiological measures were significant predictors of performance.

The NPP studies reviewed above produced significant findings, yet there are experimental flaws that influence the merit of the results for generalizing to all NPP MCR tasks. First, Lin et al. (2011) compared two subjective workload questionnaires (NASA-TLX and TWA) to determine which measure is more sensitive to task performance. However, the TWA was derived from the NASA-TLX, which confounds the results. Additionally, the additional scales seem to be measuring teamwork more than workload.

Second, Lin et al. (2010) used two different LOAs for each of the two procedures, yet compared workload of all four LOAs. In this instance, Lin and colleagues should only compare workload across all four LOA’s if the taskload of both procedures (IOP and AOP) were equal. Given that the taskload across the two procedures were unequal, the results are problematic because the varying levels of task demand across the two procedures might be driving the change in mental demand across LOAs, not the LOAs themselves. Furthermore, the data was analyzed via a between subjects Analysis of Variance (ANOVA) with a small sample size (n = 20), which is considerably below the n suggested by a power analysis.

Third, Hwang et al. (2008) used secondary task performance and seven psychological measures to investigate performance of diagnosis and monitoring tasks in the MCR of an NPP with a sample size of 15. Based on the general rules of regression, using a sample size of 15 with seven predictors can cause type I error. Most importantly, the performance measure used in the model was secondary task (mental arithmetic task) performance, not primary task (shutting down reactor) performance. Therefore, the model is predicting performance on a mental arithmetic task, not performance on shutting down a reactor. Additionally, no correlations were provided between the physiological measures and the subjective workload measures.

While it is safe to assume a change in the type of task required, during an EOP, invokes variations in workload experienced, assessing those workload changes is a more challenging task than might appear at first glance. The studies reviewed above attempted to investigate operator workload in the NPP domain, but had many flaws. Furthermore, the sensitivity of the subjective and physiological measures used in the studies is unknown. The subjective measure used by Liu et al. (2011) and Hwang et al. (2008) was the NASA-TLX, but the MRQ and ISA might be more sensitive to changes in workload in the NPP domain. The objective measures used by Hwang et al. (2008) were ECG and eye tracking, but EEG, TCD of fNIR might be more sensitive to changes in workload in the NPP domain. Identifying the workload measures that are sensitive to task type (checking, response implementation, and monitoring) changes in the NPP domain is a foundational step in NPP research. In addition to taking the findings from Abich (2013) and testing whether those results hold true in the nuclear domain, this research will guide future research in the NPP domain by identifying the workload measures that are sensitive to changes in workload in common NPP MCR tasks.

Similarities across Complex Domains.

Structure.

Workload measures might be generalizable across complex domains because these domains share many similarities. Domains such as medicine, aviation, military, and nuclear can be characterized as similar structures involving an operator or a team of operators/personnel functioning under routine conditions for a period of time before ultimately being confronted by an abnormal or emergency event that requires rapid problem solving. In other words, after a period of prolonged underloaded work operators are required to perform critical tasks, often requiring high taskload (Huey & Wickens, 1993). For example, in medical operations, when emergency medical technicians rush seriously injured patients into a hospital, emergency room personnel must move quickly to problem solve and coordinate responsibilities (Huey & Wickens, 1993). In aviation operations, after several hours without a conflict, an air traffic controller is alerted to two planes within five nautical miles laterally and 1000 feet vertically of each other and must resolve the collision course (Metzger & Parasuraman, 2005). In military operations, a Soldier conducting an intelligence, surveillance, and reconnaissance (ISR) mission suddenly notices increased enemy activity in a combat zone and must begin tracking that development (Erwin, 2013). In NPP operations, after a routine morning, an alarm alerts the team of NPP Reactor Operators (RO) of an abnormal event, such as a loss of all alternating current power to the plant's safety buses. In response, the team must establish the appropriate procedures to maintain and restore plant safety.

Factors that Drive Workload.

Complex domains share other similarities, such as factors that drive workload and influence task demand. Huey and Wickens (1993) identified several factors that drive workload and influence task demand. These factors fall into the following categories: task structure, task requirements and procedures, method that information is presented (input variables), cognitive information processing needed (information processing variables), and characteristics of response devices (output variables and computer aided and automation). The domains mentioned above have similar factors that drive workload and influence task demand, specifically in the categories of task structure, input variables, and information processing variables.

Task Structure.

Task structure has the following sub categories: performance criteria and strategies, task schedule, rate of presentation, complexity of task demands, variability of task demands, and task duration. Domains, such as military and nuclear share task structure similarities, specifically in the performance criteria and strategies, and task duration. For example, poor performance by a

Soldier tracking enemy threats can result in harm to themselves and fellow Soldiers. Similarly, poor performance by an RO can result in harm to themselves and the public. As a result, failure is typically not an option, resulting in high performance criteria and high workload (Yeh and Wickens, 1988).

Performance strategies are directly associated with mandated operating procedures or training. These procedures guide personnel throughout the task. For example, in the military, Soldiers are trained on their procedures via the crawl, walk, run method. In NPPs, ROs have to follow an Emergency Operating Procedure (EOP).

The task duration in these domains, require personnel to work long hours. Soldiers can work anywhere from eight hour to twelve hour shifts. In NPPs, many ROs work 12-hour-per day schedules (Baker, Campbell, Linder, & Moore-Ede, 1990). Long work hours lead to fatigue, which is directly associated to the workload of sustained attention (Hancock & Verwey, 1997).

Input Variables.

The method through which the information is presented (input variables) has the following sub categories: information from visual displays, information from the visual scene, and information from auditory displays. In the real world, successful task completion in complex domains requires operators to acquire information from the visual scene, whether it is from a computer display, cockpit display, or war theater. Likewise, operators often rely on information from auditory displays, such as alarms and warning tones (Huey & Wickens, 1993).

Information Processing Variables.

The cognitive information processing (information processing variables) has the following sub categories: level of processing, processing resources, memory requirements, and display control compatibility. Task performance in military operations and NPP operations share an analogous level of processing. Level of processing is operationally defined as the level and depth of analysis needed to understand and recall information. Various factors affect this level of processing and workload is linked to the amount of effort of processing required (Huey & Wickens, 1993). With this in mind, Rasmussen (1983) identified three groups of tasks (skill-based, rule-based, and knowledge-based) that vary in the level of processing required to perform each. The amount of processing required for each task is linked to the operator's level of familiarity with that task.

Skill-based tasks, which elicit the lowest workload demand, involve frequently practiced perceptual-motor skills. These tasks provide a clear relationship between the system states and the correct response (Huey & Wickens, 1993). In the nuclear domain, an example of a skill-based task is an RO pointing to a specific control on a panel. Rule-based tasks, which elicit a moderate workload demand, require the operator to perform a series of goal-oriented steps in a familiar work environment. In the NPP domain, an example of a ruled-based task is an RO in the MCR of an NPP following an EOP by shutting a valve. Knowledge-based tasks, which elicit the highest workload demand, have a high-level of unfamiliarity that requires the operator to produce new solutions. In the nuclear domain, RO's rarely perform knowledge-based tasks because all tasks are driven by procedures.

Most tasks performed by personnel in complex domains, such as military and nuclear, are rule-based tasks. Domain specific trainings and standard procedures (steps operators must follow when completing a task) accompany rule-based tasks. Typically, the goals for rule-based tasks develop through the procedures that accompany the task. For example, RO's are often directed from one EOP to another when resolving an abnormal event. Examples of rule-based tasks in the military domain are ones performed during an ISR mission, such as threat detection and change detection.

Perceptual and Physiological Processing.

When taking a closer look at the tasks performed in the military (ISR) and nuclear (NPP MCR) domains there are other similarities' besides the factors that drive workload and task structure mentioned earlier. The NPP MCR tasks included in the present experiment are checking, monitoring, and response implementation. The ISR tasks from Abich (2013) included threat detection, change detection, and multitasking. All of these tasks require perceptual and physiological processing due to the attention to stimuli and fine motor responses. The amount of stimuli processing and fine motor response varies across all three NPP MCR tasks, but both are clearly required as a part of the two-step process when completing a task. The first part of all three NPP MCR tasks require the operator to locate a specific control (visual processing of stimuli) and the second part of all three NPP MCR tasks require some type of fine motor response, such as pointing to a control or manipulating a control. The same can be said for ISR tasks. In Abich (2013), the threat detection task requires the operator to monitor (visual processing of stimuli) and report potential threats via a mouse click (fine motor response). The change detection task requires operators to monitor (visual processing of stimuli) an aerial map and identify three types of changes by using the mouse to click on a specific button (fine motor response).

Research Objective

Abich (2013) found certain subjective (ISA and MRQ) and objective measures of workload (ECG and EEG) to be universal to those changes in taskload across theoretically different tasks, but these tasks exist in the same domain. If complex domains share similarities, such as factors that drive workload, it is reasonable to assume that measures that are sensitive to workload changes in the military domain can be found to be sufficiently effective at indicating workload changes in the nuclear domain. The goal for this experiment is to determine if certain workload measures are sufficiently effective across domains by taking the findings from one domain and testing whether those results hold true in a different domain

The present experiment expands upon the findings of Abich (2013), which showed certain workload measures to be sensitive to changes in taskload within each of three different task types (change detection, threat detection, and combined). The taskload of the three tasks (checking, monitoring, and response implementation) in the present experiment will not be manipulated because the taskload level of each task is unknown. However, the nature of each task should produce different taskloads. Figure 1 illustrates the structure of each domain and highlights where comparisons were made. More specifically, the present experiment tests if the workload measures identified by Abich (2013) found to be universal and sensitive to *changes in taskload and task type* in the military domain are sufficiently effective at indicating *changes in task type* in the nuclear domain.

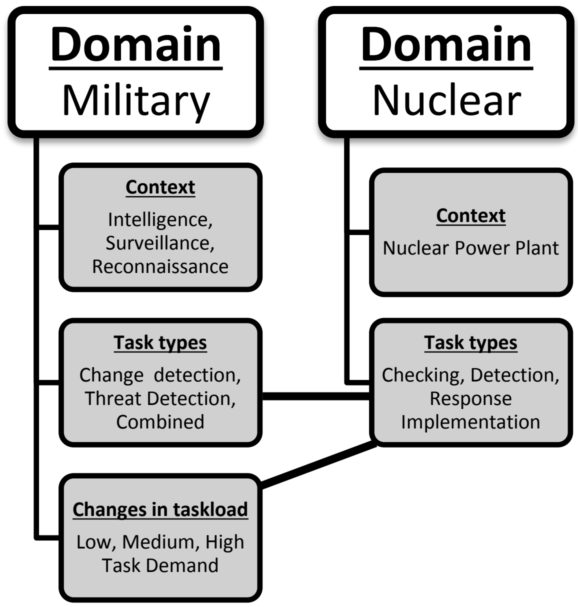


Figure 1. Structure of each domain. Comparisons were made between the changes in taskload and task type in the military domain and task type in the nuclear domain

Sufficiency Standard

The goal of this research effort is to determine if certain workload measures are *sufficiently effective* across domains by taking the findings from one domain and testing whether those results hold true in a different domain, that of nuclear.

Rose (1981) identifies three levels of translation between variables; necessary, sufficient, and exclusive. When the relationship between two variables is necessary, a relationship is present. A sufficient relationship between two variables results in a transitive relationship. An exclusive relationship means neither variable explains another phenomenon of the same general glass. Research in the field of human factors, seldom achieves exclusivity. As a result, this research effort focuses on the *sufficiency* of the identified workload measures.

Criteria from O'Donnell and Eggemeier (1986) will aid in determining what constitutes a *sufficient effective* measure of workload. O'Donnell and Eggemeier (1986) identified five criteria all workload measures should follow. All measures should; detect changes in task difficulty (sensitivity), identify the cause of these changes (diagnosticity), assess the factors related to workload (selectivity), not obstruct task performance (obtrusiveness), and consistently measure workload (reliability). For the purposes of the present study, the focus will be on sensitivity of workload measures. Specifically, the present experiment will be determine a measure is *sufficiently effective* if the workload measure detects changes in difficulty across the three identified tasks. That statement, however, is general and therefore, standards need to be applied to systematically evaluate workload measures as being *sufficiently effective.* These standards include a statistically significant *p* valve of equal to or less than .05 and an effect size identified a priori. All statistically significant workload measures with an effect size as determined by Eta squared (η^2; Cohen 1973, 1988; Pearson, 1911) of 0.138 or greater will be deemed *sufficiently effective*. According to Cohen (1973), this effect size signifies a large effect size. This approach will lead to the development of a matrix, where workload measures can fall within three areas: not significant, significant but not sufficiently effective, and sufficiently effective.

All physiological sensors yield multiple workload measures, thus each of these measures will be tested to see if they meet the standards identified. Likewise, certain workload metrics (MRQ and NASA-TLX) yield multiple workload measures, thus each will be tested to see if they meet the standards identified.

METHODOLOGY

Participants

Participants for this experiment included both undergraduate and graduate students from the University of Central Florida. Eighty-one (45 males, 36 females, $M = 21$, $SD = 4.11$) participants were recruited using an online participant pool. Participants were required to have normal or corrected-to-normal vision (including not being colorblind), and have no prior experience using a NPP simulator or operating a power plant. They were also required to have not ingested nicotine at least two hours prior to the experiment or alcohol and/or sedative medications at least 24 hours prior to the experiment.

Equipment

A customized simulator, called the Experimental Platform for Instrumentation and Control (EPIC), was utilized in the present experiment. The simulator includes one standard desktop computer (6.4GT/s, Intel Xeon™ 5600 series processor), two 24" (16:10 aspect radio) monitors, one sound bar speaker, and a customized software program called Panel Viewer. The experimental scenario consisted of tasks from common steps required when completing EOPs. EOP-EPP-001 was the foundation for the simulator's initial condition for creating the experimental scenario. However, to maintain experimental control, other realistic tasks provided by a Subject Matter Expert (SME) were incorporated. Details of the modified EOP for experimentation will become clearer in the following text.

The modified EOP provided a narrative or context by which participants operated. Specifically, it required participants to perform predetermined tasks to respond to a loss of all alternating current power to the plant's safety buses (GSE Power Systems, 2011). The modified EOP required participants to utilize two control panels (C1, A2) instead of four that would be required for to execute the full EOP associated with EOP-EPP-001. To create a modified EOP to use with participants of a novice population, the researcher made certain modifications to the EOP and the Panel Viewer panels including: reducing the amount of controls within each panel, adding additional tasks, and changing the naming convention of specific gauges and switches (Reinerman-Jones, Guznov, Mercado, D'Agostino, 2013).

The modified EOP and accompanying panels included the reduction of the amount of controls used in each panel. The first step to this method was to select the panel used in EOP-EPP-001 with the lowest amount of controls – in this case, panel C1. Next, the researcher systematically reduced the amount of controls on the A2 panel to equal the amount of controls on panel C1, which had 113 controls. The researcher calculated the reduction percentage needed to decrease the amount of controls in the A2 panel to equal the 113 controls present on panel C1. Next, the researcher categorized the controls in each panel into five groups: gauges, switches, light boxes, status boxes, and other controls. Gauges, switches, light boxes, and status boxes are the primary controls in a panel. For this experiment, participants interacted with gauges, switches, and light boxes. The researcher then reduced each type of control by the previously calculated percentage, thus leaving the ratio of control types the same on each panel. This systematic approach ensured the complexity of the original panel remains the same in the modified panel by reducing the ratio of gauges, switches, light boxes, trip boxes, and other controls. In other words, the ratio of controls on the modified panel remained intact to those of the original panel. Table 1 shows the modifications to the A2 panel. Figure 2 illustrates the original and modified A2 panels. The reduction of the amount of controls in panel A2 to equal the amount of controls in panel C1 balanced complexity across all panels, thereby removing potential confounds.

Table 1.

A2 Panel modification calculation

	Original Panel		Modified Panel	
Controls	Number of specific controls	Percent reduction needed	Calculated reduction of specific controls	Number of specific controls
		-43%		
Number of gauges	108		61.95	62
Number of switches	80		45.89	46
Number of light boxes	4		2.29	2
Number of status boxes	0		0	0
Other controls	5		2.87	3
Number of total controls	197		113	113

Figure 2. Original A2 panel used by operators (left) and modified A2 for experimentation.

In the real world, ROs refer to certain gauges and switches by their full name (e.g., moisture separator reheated bypass shut off valve). However, the names of those gauges and switches on the panels contain acronyms (e.g., MSR BYP SHUT OFF). Thus, participants would need to know the acronyms of those gauges and switches to locate them. This task would require additional training that is outside the scope of this experiment. Therefore, the researcher modified the naming convention of gauges and switches that contained both an alphanumeric code and name to decrease the difficulty of the modified EOP. Modifying the naming convention was a two-step process. First, SROs were required to refer to all gauges and switches by their alphanumeric code (i.e., STM HEADER PRESS gauge was gauge PI-464A1). Second, the researcher recoded all of the gauges and switches that have an alphanumeric code of greater than seven to an alphanumeric code of seven or less (i.e., gauge number EI-6963A1 SA was recoded to EI-6963), adhering to Miller's rule of seven plus or minus two items (Miller, 1956). Controls that do not originally have a code remained unchanged. Figures 3 and 4 illustrate the naming convention modifications.

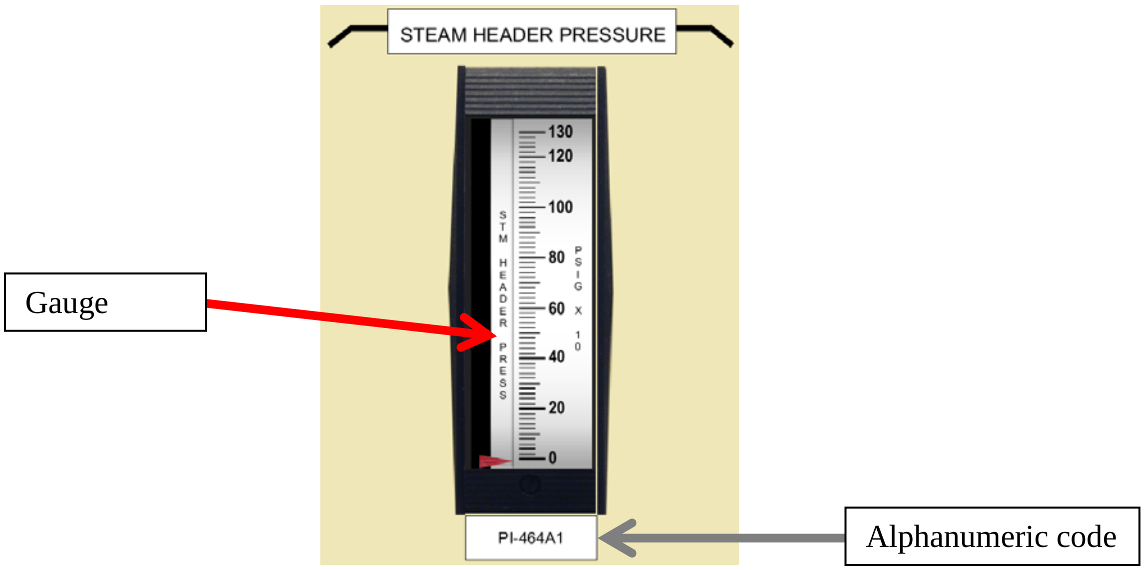


Figure 3. The red arrow on the left points to the gauge name that was not used. The grey arrow on the right points to the alphanumeric gauge code that was used.

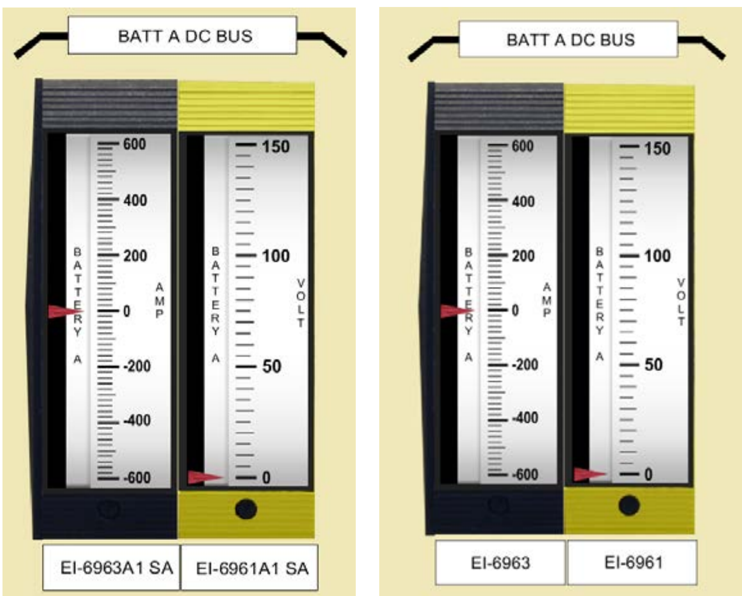


Figure 4. The two gauges shown above on the left illustrate the original gauges with an alphanumeric code of greater than seven. The two gauges shown above on the right illustrate modified gauges with an alphanumeric code of seven or less

Experimental Design

A one way repeated measures design with three levels (task type) was employed in the present experiment. There was twelve steps in each experimental scenario, grouped by task type (4 checking steps, 4 detection steps, and 4 response implementation steps). To address asymmetric transfer effects, the task types were partially counterbalanced across individual participant presentation. The task types were only partially counterbalanced to create scenarios because the tasks of checking and response implementation are directly linked such that checking always occurs before response implementation in a real NPP and thus, to maintain external validity, task yoking was observed. Scenarios were randomized and counterbalanced across participants (see Table 2). Similarly, certain steps within each task type occur in a given order due to the physics of an NPP. As a result, to ensure ecological validity, the steps within each task type were the same across participants.

Table 2.

Partial counterbalance of task types for scenario generation.

Scenario 1	Checking	Response Implementation	Detection
Scenario 2	Detection	Checking	Response Implementation
Scenario 3	Checking	Detection	Response Implementation

Independent Variables

The independent variable in this experiment were task type (i.e., checking, detection, and response implementation).

Task Type.

Task type consisted of three conditions. The checking task type required a one-time inspection of an instrument or control to verify that it was in the state that the EOP calls for it to be. Participants were required to locate light boxes and valves and indicate identification by clicking on the correct control. The detection task type required participants to correctly locate a control then continuously monitor that control parameter for identification of change. Participants were required to monitor the gauge for five minutes and detect changes in level by clicking on an acknowledge button located at the bottom of the display. Twelve random changes per minute occurred, totaling 60 changes per detection task. The response implementation task type required an action to affect the state of the NPP. Similar to the checking and detection steps, participants were required to correctly identify a control, then open or shut a switch on that control. Each task type consisted of four steps that were executed using three-way communication led by the experimenter acting as the SRO.

Dependent Variables.

Demographics.

A demographics questionnaire was used to gather information about age, sex, educational level, computer and television usage, and hours of sleep the night before participation.

Performance

Execution Performance.

Checking.

Correct identification of controls and erroneous identifications were recorded by the simulator. A Kinect with Microsoft Voice Recorder recorded verbal verification of the checked light box or valve.

Detection.

The EPIC simulator recorded hits, misses, and false alarms. A Kinect with Microsoft Voice Recorder recorded verbal verification when the gauge level reaches the specified amount.

Response Implementation.

The EPIC simulator recorded correct and incorrect actions. A Kinect with Microsoft Voice Recorder recorded verbal verification of the completed action.

Communication (Instruction) Performance.

A Kinect with Microsoft Voice Recorder recorded verbal three-way communication. Three-way communication performance measures included instruction events per task, instruction events repeated, instruction clarifications, location help, and percent correct. Instruction events per task were the number of three-way communication events completed. An instruction event repeated was the number of requests by participants for a repeated instruction and the number of request by the SRO for a repeated response from participants. An instruction clarification was a clarification by the SRO to a participant. Location help was the number of requests, by participants, for assistance in locating the correct control. Percent correct was the percentage of correct responses, on all six parts of three-way instruction.

Subjective Measures.

NASA- TLX.

The TLX (Hart & Staveland, 1988, 2006) questionnaire was used to assess each participant's perceived workload using a multi-dimensional scale with subscales. The subscales include mental demand, physical demand, temporal demand, effort, frustration, and performance. The TLX uses a 100-point sliding scale with five-point increments to rate each subscale. The average score of the six subscales provided a separate measure of global workload. Participants received a copy of the scale with subscale definitions and completed the TLX at the end of each task type throughout the scenario using a customized computer program that automatically activated a visual prompt containing the questionnaire.

ISA.

The ISA (Hulbert, 1989; Jordan, 1992) was used to measure immediate subjective workload assessed during the performance of a task, using a five-point Likert scale (Tattersall & Foord, 1996). Participants received a copy of the scale with definitions and complete the ISA halfway through each task type using a customized computer program that automatically activated an audio prompt containing the questionnaire. The audio prompt contained the phrase, "please rate your workload," signaling participants to respond by writing down their rating on a sheet of paper.

MRQ.

The MRQ was used to characterize the nature of the mental processes used during each task (Boles & Adair, 2001). The items on the questionnaire are derived from factor analytic studies of lateralized processes (Boles, 1991, 1992, 1996, 2002). Participants received a copy of the scales, with definitions, and complete the MRQ at the end of each task type throughout the scenario using a customized computer program that automatically activated a visual prompt containing the questionnaire. Boles (1996) indicates that the MRQ is most effective when only the target scales for the task are included. The following 14 of 17 scales were included for the present experiment: auditory emotional process, auditory linguistic process, manual process, short-term memory process, spatial attentive process, spatial categorical process, spatial concentrative process, spatial emergent process, spatial positional process, spatial quantitative process, visual lexical process, visual phonetic process, visual temporal process, and vocal process. Ten of the 14 scales are the same as Abich (2013). The present experiment included four additional scales (auditory emotional process, spatial categorical process, spatial quantitative process, and visual phonetic process) to allow assessment of all aspects of the NPP MCR tasking environment to ensure fair evaluation of the MRQ's utility as a measure of workload in the NPP domain.

Physiological Measures

Electroencephalogram (EEG).

The Advanced Brain Monitoring B-Alert X10 system was employed to assess nine-channels of EEG and one channel of ECG (Figure 5). Following the international standard 10-20

System, the sampling rate of 256 Hz captured signals from Fz, F3, F4, Cz, C3, C4, Pz, P3, and P4. Reference electrodes were placed on each participant's mastoid bone. PSD analysis techniques were used to analyze three standard bandwidths: theta (4-8 Hz), alpha (9-13 Hz), and beta 14-30 Hz (Wilson, 2002). Each bandwidth was collected for the nine nodes. They were combined to compare left and right hemispheres and the frontal, temporal, and parietal lobes.

Figure 5. ABM's X10 EEG/ECG system

Transcranial Doppler (TCD).

The Spencer Technologies' ST^3 Digital Transcranial Doppler, model PMD150, was used to monitor CBFV of the medial cerebral artery (MCA) in the left and right hemisphere through high pulse repetition frequency (PRF; Figure 6). The Marc 600 head frame set was used to hold the TCD probes in place.

Figure 6. Spencer Technologies' ST[3]
Transcranial Doppler

functional Near Infra-Red Imaging (fNIR).

The Somantics' Invos Cerebral/Somatic Oximeter, model 5100C, was used to measure (hemodynamic) changes in oxygenated hemoglobin (oxy-Hb) and deoxygenated hemoglobin (deoxy-HB) in the prefrontal cortex (Ayaz et al., 2011; Chance, Zhuange, UnAh, Alter, & Lipton, 1993; Figure 7).

Figure 7. fNIR strip

Electrocardiogram (ECG).

The Advanced Brain Monitoring System B-Alert X10 system was used to monitor the ECG, sampling at 256 Hz. Single-lead electrodes were placed on the center of the right clavicle and one on the lowest left rib (Figure 8). Heart Rate (HR) was computed using peak cardiac activity to measure the interval from each beat per second. "So and Chan" QRS detection methods was used to calculate IBI and HRV (Taylor, Reinerman-Jones, Cosenzo, & Nicholson, 2012). This approach maximizes the amplitude of the R-wave (Henelius, et al., 2009).

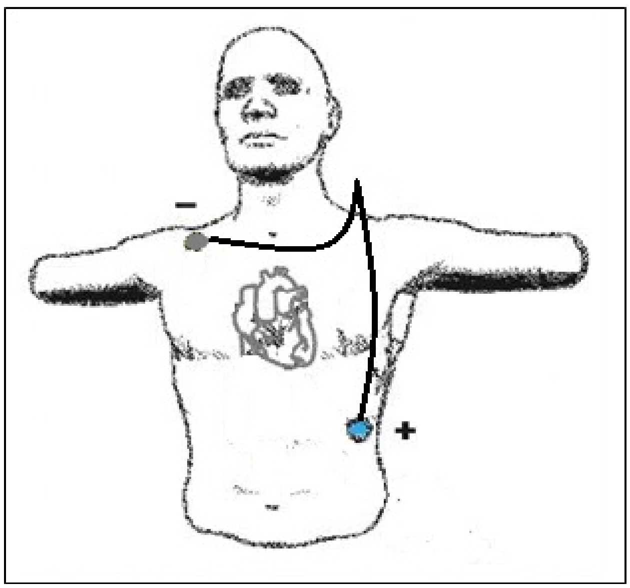


Figure 8. Electrode locations for the ECG system

Procedure

Participants were provided with a copy of the informed consent, followed by the Ishihara color-blind test and the demographics questionnaire. Participants were then train for two hours using a PowerPoint presentation and the EPIC simulator. The presentation provided an introduction to the procedures and protocols for participating in a NPP simulation for experimental research. Participants were trained to use 3-way communication to clearly relay critical information, navigate within the EPIC simulator to locate and read status indicators, respond appropriately to a simulated NPP system warning by following standardized procedures, and complete questionnaires. Each aspect was trained separately and then a practice session combined all components. Feedback and proficiency tests were given after each portion. Participants' scores had to be over 80% to move forward to the experimental scenario. After training, participants were given a five-minute break. The physiological sensors were connected and a five-minute resting baseline was taken before proceeding with the first task type of the experimental scenario. The steps within the task type were carried-out implementing three-way communication protocol initiated by the experimenter acting as the SRO. The ISA rating was prompted halfway through the condition and the TLX and MRQ were administered after each task condition block. The same process was followed for the next two task type conditions. The experimental session finished by disconnecting the physiological sensors. Experimental sessions were two hours.

Experimental Hypothesis

Hypothesis 1.

It is hypothesized that both subjective and objective measures will be sufficiently effective and indicating changes in workload associated with task performance in NPP MCR operations.

Hypothesis 2.

In regards to subjective workload measures, ISA will provide a global online assessment of workload associated with task types in a complex NPP MCR operation. When comparing the TLX and MRQ, the MRQ will be superior because it will detect dimensions affecting workload that the TLX cannot identify for each of the task types.

Hypothesis 3.

In regards to objective workload measures, ECG, specifically HRV, will prove to be the most promising physiological measure of workload differences associated with task types in a complex NPP MCR operation.

RESULTS

Analysis

Analyses of Variance (ANOVAs) was used to conduct a task demand check to determine if each task type yielded distinct levels of workload as assessed by subjective and objective measures. The Greenhouse-Geisser correction in SPSS was applied to correct for violations of sphericity. Bonferroni corrections were used to post-hoc comparisons to account for the chance of Type I errors. Effect sizes, means, and standard deviations were reported. Correlations between subjective and physiological measures were conducted. Multiple regression analysis was used to show how well the subjective and physiological measures could predict overall performance on each task type. For ease of use by the reader, only significant results were graphed and tabled unless otherwise specified.

Effectiveness Checks

Subjective Measures.

NASA-TLX.

A 3 (checking, detection, response implementation) × 6 (mental demand, temporal demand, physical demand, effort, performance, frustration) repeated measures ANOVA was run to determine if workload was significantly different between the task types, if the type of workload was different across task types, and if the type of workload was different for each of the task types.

Results indicate a statistically significant main effect for task type $F(2, 160) = 4.038$, $p = .019$, $\eta^2 = .013$, such that the detection task type ($M = 38.85$) was significantly more demanding overall than the response implementation task type ($M = 34.02$). A significant main effect was found for the NASA-TLX $F(3.446, 275.647) = 31.711$, $p < .000$, $\eta^2 = .284$, such that mental demand ($M = 46.66$) was greater than the other five subscales. The interaction effect was statistically significant $F(6.930, 554.371) = 10.669$, $p < .000$, $\eta^2 = .118$.

Six one-way repeated measures ANOVAs with three levels (checking, detection, response implementation) were conducted to identify the type of demand per task type. Results indicate that task type had a significant effect on physical demand, $F(1.847, 147.783) = 10.804$, $p < .000$, $\eta^2 = .020$, such that both the detection ($M = 20.59$) and response implementation ($M =17.31$) task types were significantly more physically demanding than the checking task type ($M = 12.84$). A significant effect was found for temporal demand, $F(1.715, 137.227) = 4.107$, $p = .024$, $\eta^2 = .018$, such that the checking task type ($M = 43.54$) was significantly more temporally demanding than the detection task type ($M = 34.99$). A significant effect was found for frustration, $F(1.749, 139.936) = 34.069$, $p < .000$, $\eta^2 = .138$, such that the detection task type ($M = 51.26$) was significantly more frustrating than both the checking ($M = 29.14$) and response implementation task types ($M = 26.73$, see figure 9 and table 3). There were no significant task type differences for mental demand, effort, and performance.

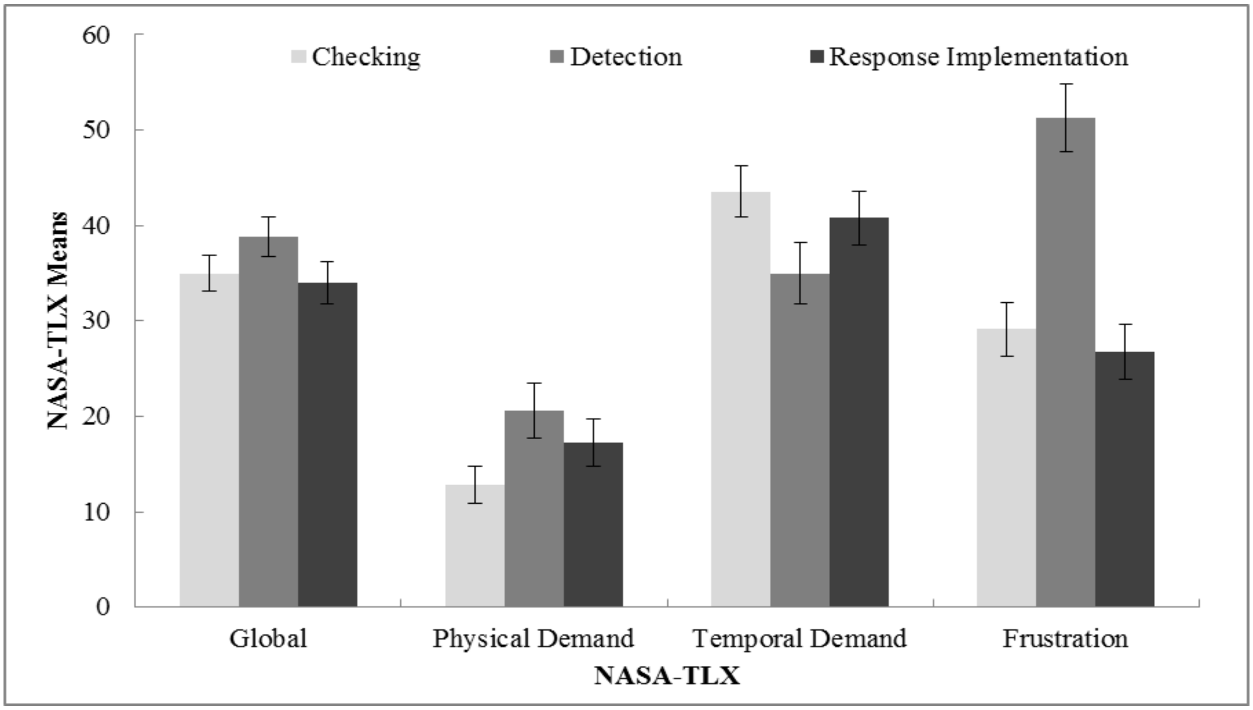


Figure 9. NASA-TLX ratings. Error bars in this figure represent standard errors.

Table 3

NASA-TLX means and standard deviations (in parentheses) for each task type

NASA-TLX Variables	Checking	Detection	Response Implementation
Global	34.99 (16.97)	38.85 (18.90)	34.02 (19.53)
Physical demand	12.84 (18.01)	20.59 (25.64)	17.31 (22.22)
Temporal demand	43.54 (22.95)	34.99 (28.99)	40.78 (25.56)
Frustration	29.14 (25.45)	51.26 (31.87)	26.73 (25.65)

MRQ.

MRQ results (14 scales detailed in the Method) for all three task types (checking, detection, response implementation) were analyzed via one-way repeated measures ANOVAs to determine if there was a significant difference between task type for each scale. MRQ results indicate that task type had a significant effect on spatial concentrative, $F(2, 160) = 5.330$, $p = .006$, $\eta^2 = .026$, such that that the detection task type ($M = 64.31$) required significantly more spatial concentrative processing than both the checking (*M* = 54.90) and response implementation (*M* = 56.49) task types. A significant effect was found for visual temporal, $F(2, 160) = 8.978$, $p < .000$, $\eta^2 = .032$, such that the detection task type (*M* = 55.99) required significantly more visual temporal processing than both the checking (*M* = 44.09) and response implementation task type (*M* = 46.36). A significant effect was found for spatial quantitative, $F(2, 160) = 7.013$, $p = .001$, $\eta^2 = .028$, such that the detection task type (*M* = 63.79) required significantly more spatial quantitative processing than both the checking (*M* = 54.95) and response implementation (*M* = 53.89) task types. A significant effect was found for spatial attentive, $F(2, 160) = 3.875$, $p = .023$, $\eta^2 = .017$, such that the detection task type (*M* = 78.52) required significantly more spatial attentive processing than the response implementation (*M* = 73.00) task type. A significant effect was found for spatial positional, $F(1.860, 148.838) = 3.989$, $p = .023$, $\eta^2 = .021$, such that checking task type (*M* = 73.04) required significantly more spatial positional processing than the response implementation task type (*M* = 66.24). A significant effect was found for vocal process, $F(2, 160) = 4.896$, $p = .009$, $\eta^2 = .009$, such that the response implementation task type (*M* = 67.96) required significantly more vocal processing than the detection task type (*M* = 62.06, see figure 10 and table 5). No other pairwise comparisons reached such a significant level of distinction. There were no significant task type differences for auditory emotional, auditory linguistic, manual process, short term memory, spatial categorical, spatial emergent, visual lexical, and visual phonetic.

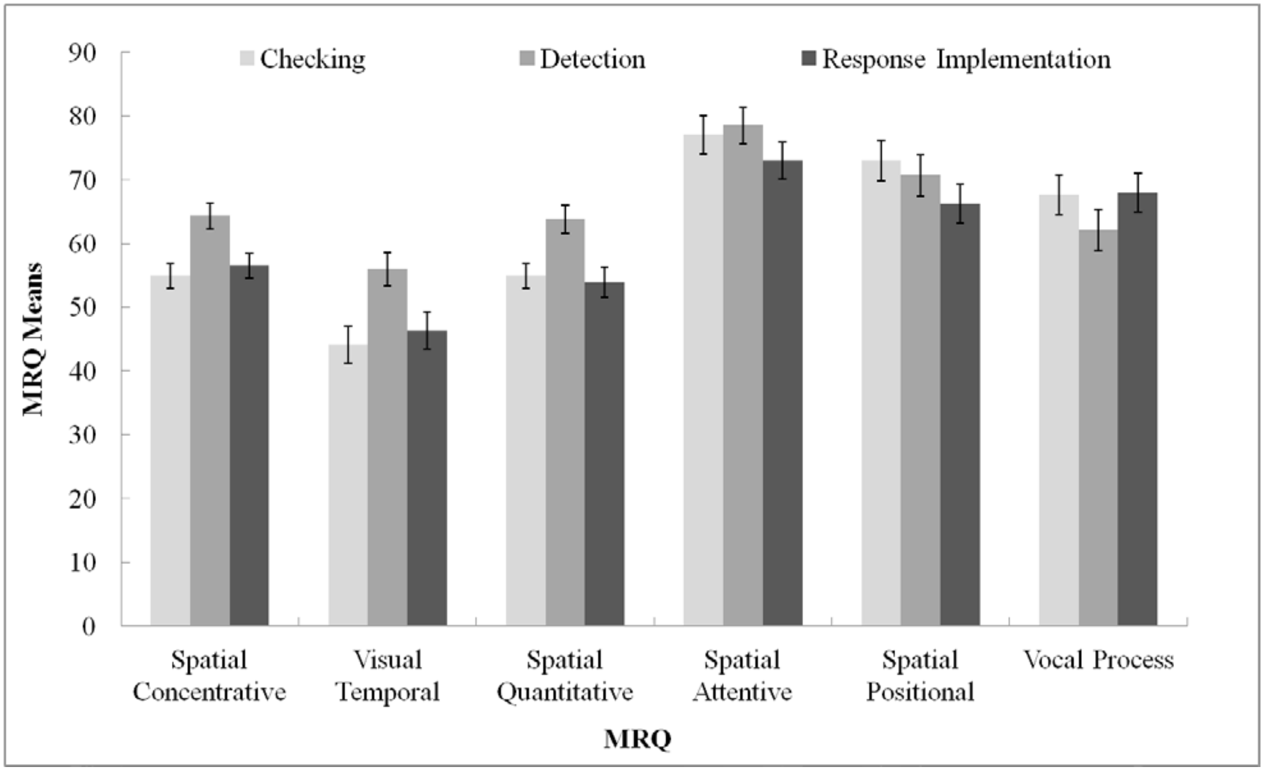


Figure 10. MRQ ratings. Error bars in this figure represent standard errors.

Table 4

MRQ means and standard deviations (in parentheses) for each task type

MRQ Variables	Checking	Detection	Response Implementation
Spatial concentrative	54.90 (25.84)	64.31 (23.30)	56.49 (26.10)
Visual temporal	44.09 (28.58)	55.99 (28.85)	46.36 (27.62)
Spatial quantitative	54.95 (26.70)	63.79 (25.41)	53.89 (26.57)
Spatial attentive	77.06 (17.54)	78.52 (18.42)	73.00 (17.28)
Spatial positional	73.04 (17.39)	70.68 (19.44)	66.24 (21.64)
Vocal process	67.65 (28.12)	62.06 (29.68)	67.96 (27.40)

ISA.

ISA results for all three task types (checking, detection, response implementation) were analyzed via a one-way repeated measures ANOVA to determine if task type had a significant effect on online-subjective workload. There were no significant findings in ISA rating between task types.

Performance Measures.

Instruction Performance.

Instruction performance measures included percent correct, location help, clarification, and request for repeat instruction. A one-way (checking, detection, response implementation) repeated measures ANOVA was run for each of those four measures to determine if there is a significant difference between task types.

Instruction performance results indicate that task type had a significant effect on percent correct $F(1.742, 139.335) = 16.974$, $p < .000$ $\eta^2 = .088$, such that percent correct for the checking ($M = 90.40$) and response implementation task types ($M = 94.16$) were significantly higher than the detection task type ($M = 82.10$, see figure 11 and table 5). A significant effect was found for clarification $F(1.462, 116.950) = 60.561$, $p < .000$ $\eta^2 = .298$, such that request for clarifications for the detection task type ($M = 1.98$) was significantly higher than both the checking ($M = .407$) and response implementation ($M = .432$) task types (see figure 12 and table 5). A significant effect was found for request for repeat instruction $F(1.308, 104.675) = 55.488$, $p < .000$ $\eta^2 = .301$, such that request for repeat instruction for the detection task type ($M = 1.43$) was significantly higher than both the checking ($M = .198$) and response implementation ($M = .247$) task types, see figure 12 and table 5). There were no significant findings for location help between task types.

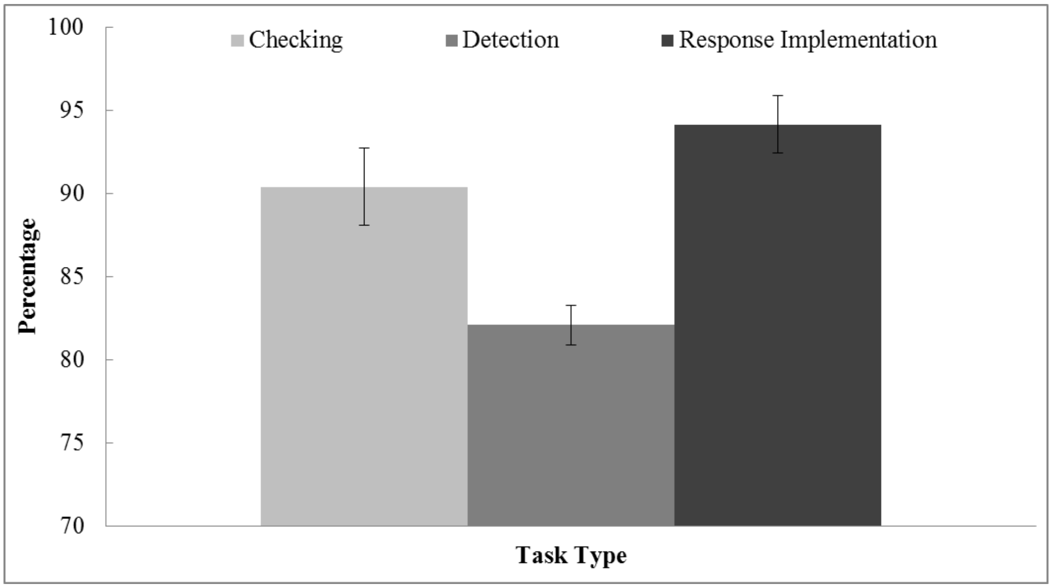


Figure 11. Instruction percent correct. Error bars in this figure represent standard errors.

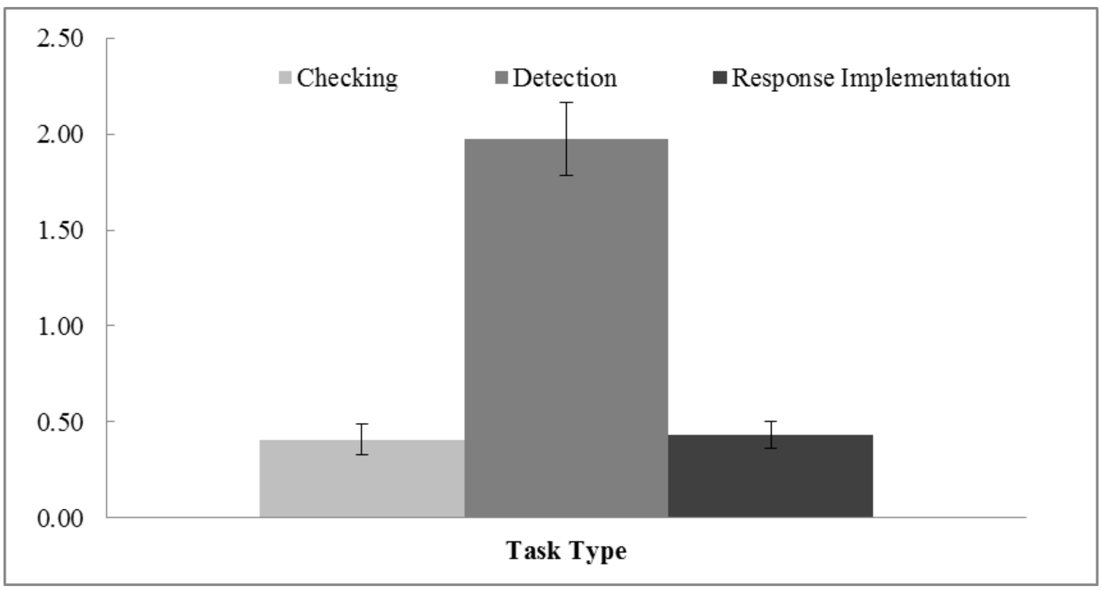


Figure 12. Instruction clarification. Error bars in this figure represent standard errors.

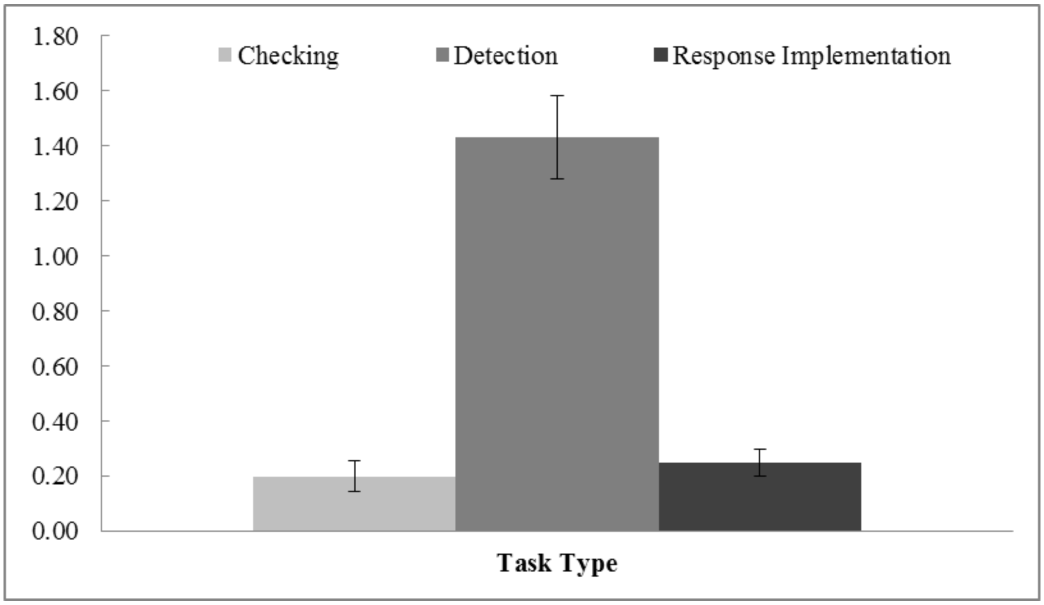


Figure 13. Request for repeat instruction. Error bars in this figure represent standard errors.

Table 5

Instruction performance means and standard deviations (in parentheses) for each task type

Instruction Performance Variables	Checking	Detection	Response Implementation
Percent correct	90.40 (20.85)	82.10 (15.67)	94.16 (10.69)
Clarification	.407 (.72)	1.98 (1.70)	.432 (.65)
Request for repeat instruction	.198 (.51)	1.43 (1.36)	.25 (.43)

Overall Performance.

The overall performance measure was an average of instruction performance percent correct and execution performance percent correct. A one-way (checking, detection, response implementation) repeated measures ANOVA was run for overall performance to determine if there is a significant difference between task types. There were no significant findings for location help between task types. Table 6 shows the mean and standard deviations for each task type.

Table 6

Overall performance means and standard deviations (in parentheses) for each task type

Variable	Checking	Detection	Response Implementation
Overall performance	75.78 (19.62)	73.63 (12.99)	77.24 (18.75)

Physiological Measures.

All dependent variables entered into the ANOVAs were the difference from a five-minute baseline. For example, if the participant's left CBFV for the five-minute baseline was 63.23 cm/s and their left CBFV for the subsequent checking task was 65.32 cm/s, their difference from baseline would be 2.09 cm/s. This approach helps account for individual differences when comparing group means as is the case when running ANOVAs.

EEG.

A 3 (checking, detection, response implementation) × 2 (left and right hemisphere difference from baseline) repeated measures ANOVA was run for Alpha, Beta, and Theta to determine if there were significant differences for each frequency band for each hemisphere between task types. A 3 (checking, detection, response implementation) × 3 (frontal, parietal, occipital lobe difference from baseline) repeated measures ANOVA was run for Alpha, Beta, and Theta to determine if there were significant differences for each frequency band for each lobe between task types. A 3 (checking, detection, response implementation) × 9 (F3, Fz, F4, C3, Cz, C4, P3, POz, P4 difference from baseline) repeated measures ANOVA was run for Alpha, Beta, and Theta to determine if there were significant differences for each frequency band for each channel between task types. There were no significant EEG findings between task types.

TCD.

A 3 (checking, detection, response implementation) × 2 (left and right hemisphere difference from baseline) repeated measures ANOVA was run to determine if CBFV was significantly different for task type, if CBFV was significantly different for hemisphere, and if CBFV for one hemisphere for one task type was significantly different than the other hemisphere and other task types. The main effect of task type was statistically significant $F(2, 152) = 4.125$, $p = .018$, $\eta^2 = .010$, such that the checking task type ($M = .290$) yielded a significantly higher CBFV difference from baseline than the response implementation task type ($M = -.543$). The main effect of CBFV hemisphere and the interaction effect were not significant.

Two one-way repeated measures ANOVAs with three levels (checking, detection, response implementation) were conducted to identify CBFV region (left and right hemisphere) differences per task type. Results indicate that task type had a significant main effect on left hemisphere, $F(2, 152) = 3.568$, $p = .031$, $\eta^2 = .008$, such that left CBFV difference from baseline for the checking task type ($M = .830$) was significantly higher than the response implementation task type ($M = .002$, see figure 14 and table 7). There were no significant findings for the right hemisphere between task types.

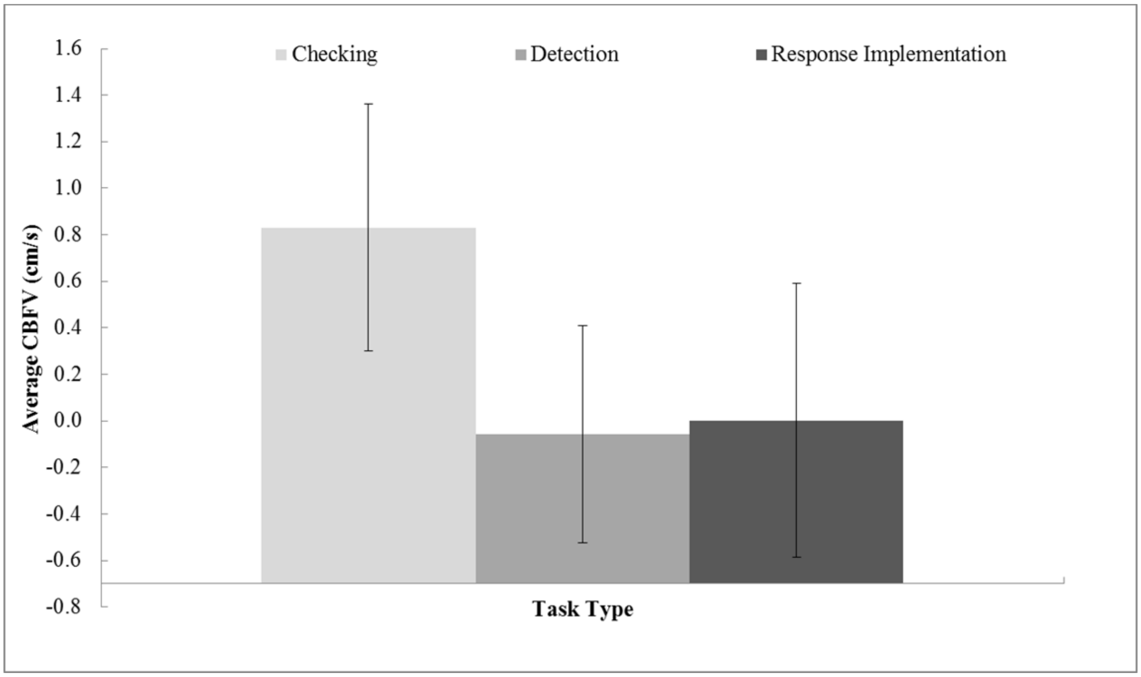


Figure 14. Left TCD CBFV difference from baseline, where 0 was the baseline. Error bars in this figure represent standard errors.

Table 7

TCD mean difference from baseline and standard deviations (in parentheses) for each task type

TCD Variable	Checking	Detection	Response Implementation
Left CBFV (cm/s)	0.83 (4.66)	-0.58 (4.10)	.002 (5.16)

fNIR.

A 3 (checking, detection, response implementation) × 2 (left and right hemisphere) repeated measures ANOVA was run to determine if oxygenation was significantly different for task type, if oxygenation was significantly different for hemisphere, and if oxygenation for one hemisphere for one task type was significantly different than the other hemisphere and other task types.

The main effect of task type was statistically significant $F(2, 150) = 17.633$, $p < .000$, $\eta^2 = .190$, such that the detection task type ($M = -.272$) yield a significantly greater blood oxygenation difference from baseline than both the checking ($M = -.831$) and response implementation ($M = -1.06$) task types. The main effect of hemispheric blood oxygenation and interaction effect were not significant.

Two one-way repeated measures ANOVAs with three levels (checking, detection, response implementation) were conducted to identify blood oxygenation (left and right hemisphere) region differences per task type. Results indicate that task type had a significant effect on left hemisphere, $F(2, 156) = 10.361$, $p < .000$, $\eta^2 = .025$, such that left frontal cortex blood oxygenation difference from baseline for the detection task type ($M = .580$) was significantly higher than the response implementation task type ($M = -1.367$). A significant effect was found for right hemisphere, $F(2, 150) = 22.701$, $p < .000$, $\eta^2 = .048$, such that right frontal cortex blood oxygenation difference from baseline for the detection task type ($M = -.650$) was significantly higher than both the checking ($M = -.743$) and response implementation ($M = -.950$) task types (see figure 15 and table 8).

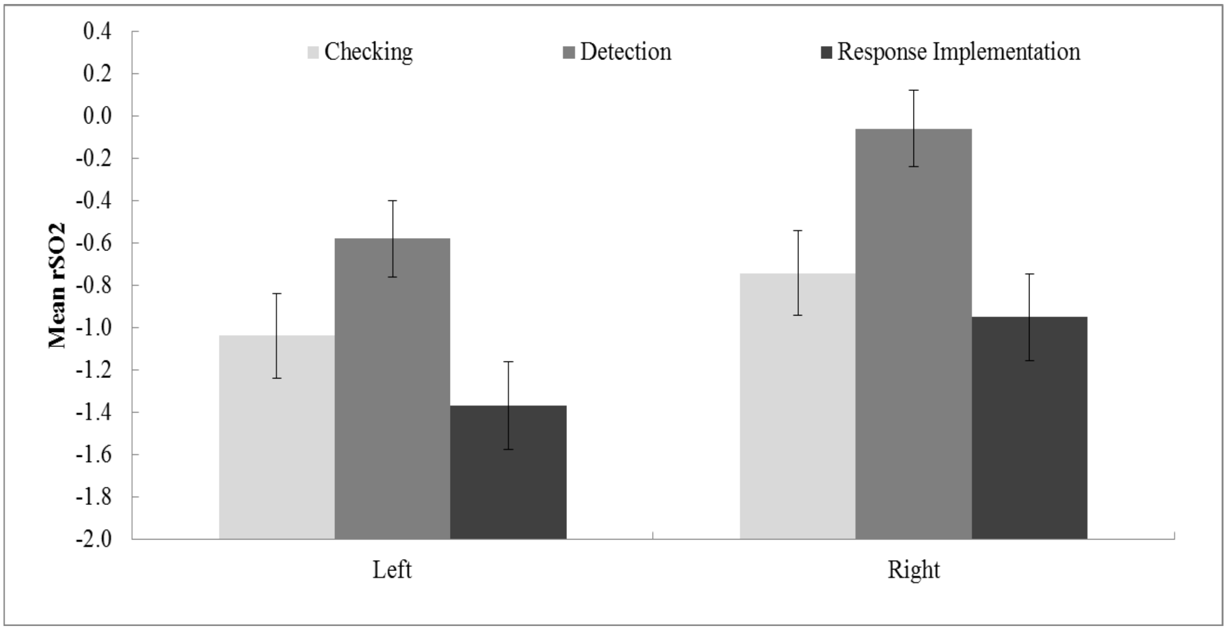


Figure 15. Frontal cortex blood oxygenation difference from baseline during all three task types, where 0 was the baseline. Error bars in this figure represent standard errors.

Table 8

fNIR mean difference from baseline and standard deviations (in parentheses) for each task type

fNIR Variable	Checking	Detection	Response Implementation
Left	-1.04 (1.81)	-0.58 (2.00)	-1.37 (2.37)
Right	-0.74 (1.75)	-0.06 (1.58)	-0.95 (1.79)

ECG.

Three one-way (checking, detection, response implementation) repeated measures ANOVAs were run for HR, HRV, and IBI to determine if heart response was significantly different between the task types.

Results indicate that task type had a significant main effect on HR, $F(1.401, 102.244) = 4.541$, $p = .024$, $\eta^2 = .022$, such that HR for the checking task type (M =3.421) was significantly higher than the detection task type (M = 1.348). A significant effect was found for IBI, $F(2, 146) = 6.422$, $p = .002$, $\eta^2 = .017$, such that IBI for the checking task type (M =58.18) was significantly higher than the detection task type (*M* = 38.72, see figures 16 and 17 and table 9). There were no significant results for HRV.

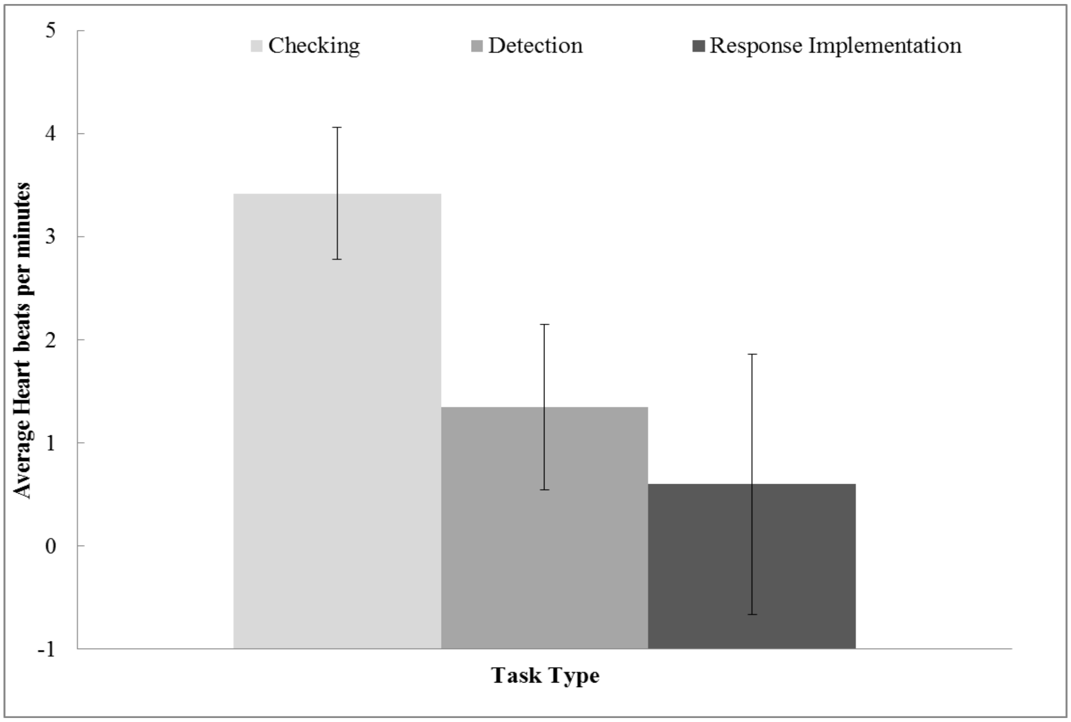


Figure 16. Average heart beats per minute difference from baseline during all three task types. Error bars in this figure represent standard errors.

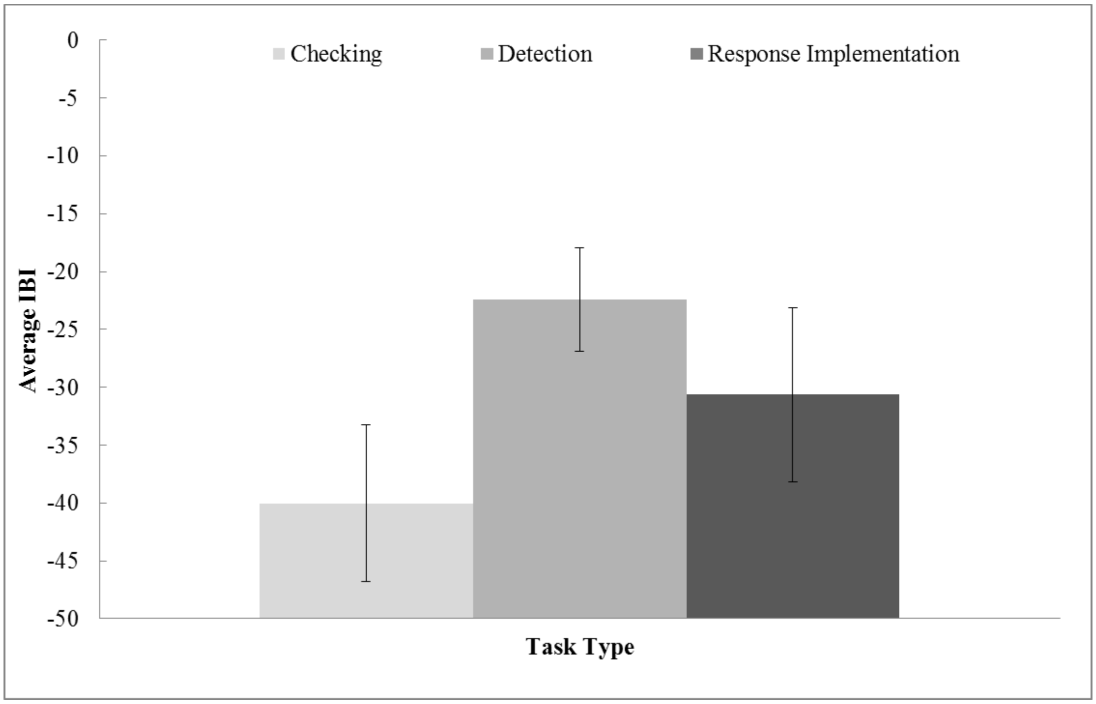


Figure 17. Average interbeat interval difference from baseline during all three task types. Error bars in this figure represent standard errors.

Table 9

ECG mean difference from baseline and standard deviations (in parentheses) for each task type

ECG Variables	Checking	Detection	Response Implementation
HR	3.421 (5.50)	1.348 (6.89)	0.6 (10.85)
IBI	-40.04 (58.18)	-22.42 (38.72)	-.30.65 (64.74)

Correlations

Pearson correlation analysis was used to assess the relationship between physiological and subjective measures of workload for all three task types.

Checking Task.

Pearson's correlation analysis revealed a positive correlation between right hemisphere CBFV and NASA-TLX frustration (r = .232, p = .043) and between left hemisphere CBFV and MRQ vocal process (r = .243, p = .032). On the other hand, correlation analysis revealed a negative correlation between right hemisphere CBFV and MRQ auditory linguistic (r = -.265, p = .020).

Results indicated a negative relationship between EEG right hemisphere, for all three waves and various subjective measures. EEG Alpha right hemisphere was negatively correlated with NASA-TLX mental demand (r = -.286, p = .011), NASA-TLX temporal demand (r = -.242, p = .031), and MRQ visual temporal (r = -.259, p = .021). EEG Beta right hemisphere was negatively correlated with NASA-TLX mental demand (r = -.286, p = .011), NASA-TLX temporal demand (r = -.245, p = .029), and MRQ visual temporal (r = -.259, p = .021). EEG Theta right hemisphere was negatively correlated with NASA-TLX mental demand (r = -.286, p = .011), NASA-TLX temporal demand (r = -.246, p = .029), and MRQ visual temporal (r = -.259, p = .021).

Results also indicated a negative relationship between HR and MRQ manual (r = -.284, p = .014), MRQ short-term memory (r = -.304, p = .008), MRQ spatial attentive (r = -.260, p = .025), and MRQ spatial emergent (r = -.259, p = .026, see table A1).

Detection Task.

Pearson's correlation analysis revealed a positive correlation between left hemisphere CBFV and MRQ manual (r = .293, p = .009). Both left hemisphere blood oxygenation (r = .264, p = .019) and right hemisphere blood oxygenation (r = .239, p = .037) were positively correlated with MRQ manual. Results also indicated a positive relationship between HR and NASA-TLX temporal demand (r = .264, p = .023).

There were several negative correlations between physiological and subjective measures. EEG Theta left hemisphere was negatively correlated with MRQ spatial quantitative (r = -.222, p = .049). Right hemisphere CBFV was negatively correlated with MRQ manual (r = -.326, p = .004). Right hemisphere blood oxygenation was negatively correlated with MRQ spatial positional (r = -.231, p = .043) and MRQ spatial quantitative (r = -.242, p = .034). In addition, results indicated a negative relationship between IBI and NASA-TLX mental demand (r = -.335, p = .004), NASA-TLX effort (r = -.329, p = .004), MRQ auditory linguistic (r = -.383, p = .001), and MRQ short-term memory (r = -.258, p = .026, see table A2).

Response Implementation Task.

Unlike the Pearson's correlations results for the checking and detection task types, correlation analysis for the response implementation task type revealed several positive correlations. EEG Alpha left hemisphere was positively correlated with MRQ auditory linguistic (r = .308, p = .006) and MRQ vocal process (r = .251, p = .025). EEG Beta left hemisphere was positively correlated with MRQ auditory linguistic (r = .309, p = .006) and MRQ vocal process (r = .252, p = .025). EEG Theta left hemisphere was positively correlated with MRQ auditory linguistic (r = .309, p = .006) and MRQ vocal process (r = .253, p = .024). Left hemisphere CBFV was positively correlated with MRQ manual (r = .270, p = .018), MRQ short-term memory (r = .229, p = .035), MRQ visual lexical (r = .306, p = .007), and MRQ vocal process (r = .431, p < .001). Left hemisphere blood oxygenation was positively correlated with NASA-TLX frustration (r = .247, p = .028) and MRQ visual lexical (r = .223, p = .048), while right hemisphere blood oxygenation was positively correlated with NASA-TLX frustration ($r = .258$, $p = .025$). Results also indicated a positive relationship between HR and MRQ spatial attentive ($r = .238$, $p = .040$).

Right hemisphere blood oxygenation was the only physiological measure negatively correlated with subjective measures. Right hemisphere blood oxygenation was negatively correlated with MRQ spatial categorical ($r = -.256$, $p = .025$), MRQ spatial quantitative ($r = -.269$, $p = .019$), and MRQ visual phonetic ($r = -.236$, $p = .040$, see table A3).

Hierarchical Regressions

Hierarchical multiple regression was used to assess the ability of subjective (NASA-TLX, ISA, MRQ) and physiological measures (EEG, TCD, fNIR, ECG) to predict performance. Pairwise deletions were applied when necessary. Preliminary analyses were conducted to ensure no violations of assumptions of normality, linearity, multicollineraity, and homoscedasticity. Based on the theoretical assumptions that subjective measures are the most widely accepted assessments of workload and a similar analysis approach used by Abich (2013), subjective measures that were entered at Step 1 included the NASA-TLX, MRQ, and ISA. To test for incremental variance accounted for by physiological measures, the variables that were be entered at Step 2 include Alpha, Beta, and Theta for each EEG channel, CBFV in the left and right hemisphere, oxygenation in the left and right hemisphere, HR, HRV and IBI.

Overall performance for a single task type (checking, detection, response implementation) was regressed on workload measures for that specific task type, making three hierarchal regressions. For example, the performance on the checking task was regressed on the checking task workload measures. Results presented in tables represent those predictors that significantly contributed to the model.

Checking Task.

Checking task subjective measures entered at Step 1, significantly relate to checking task performance, $F(21, 49) = 2.508$, $p = .004$. However, due to multicollinearity, there were no significant predictors of the model. After entry of physiological measures at Step 2, the model was not significant, $F(44, 26) = 1.62$, $p = .095$ (see table 10).

Table 10

Results of regressing checking task performance on checking task workload variables

		R	df	$R^2_{Adjusted}$	$R^2_{changed}$	*B*	*SE B*	β
Step 1		0.720	21, 49	0.311	0.518			
	(Constant)					50.720	15.479	
Step 2		0.856	23, 26	0.281	0.215			
	(Constant)					255.567	109.458	
	MRQ Vocal Process					0.215	0.104	0.347
	EEG Theta P4					-0.043	0.017	-2.020

Detection Task.

Detection task subjective measures entered at Step 1, did significantly relate to detection task performance, $F(21, 50) = 2.093$, $p = .017$. After entry of physiological measures at Step 2, the model was still significant, $F(41, 30) = 2.638$, $p = .003$ (see table 11).

Table 11

Results of regressing detection task performance on detection task workload variables

		R	df	$R^2_{Adjusted}$	$R^2_{changed}$	*B*	*SE B*	β
Step 1		0.684	21, 50	0.244	0.468			
	(Constant)					62.579	9.756	
	NASA-TLX Frustration					-0.177	0.057	-0.431
	ISA					5.724	1.711	0.427
Step 2		0.885	20, 30	0.486	0.315			
	(Constant)					6.345	46.622	
	NASA-TLX Frustration					-0.180	0.067	-0.438
	NASA-TLX Performance					0.113	0.054	0.262
	ISA					6.312	1.926	.471
	Heart Beats per minute					0.605	0.237	0.524

Response Implementation Task.

Response implementation task subjective measures entered at Step 1, was not significantly relate to response implementation task performance $F(21, 47) = 1.159$, $p = .328$. After entry of physiological measures at Step 2, the model was significant not $F(43, 25) = 1.696$, $p = .08$ (see table 12).

Sufficiency Standard

The sufficiency standard matrix developed in the introduction revealed that one type of each measure was sufficiently effective at indicating workload changes between task types. The subjective measure that was sufficiently effective was the NASA-TLX frustration and the physiological measure that was sufficiently effective was fNIR (see table 12).

Table 12

Sufficiency standard matrix

	Subjective Measures				Physiological Measures		
Variables	*Not Significant*	*Significant but Not Sufficiently Effective*	*Sufficiently Effective*		*Not Significant*	*Significant but Not Sufficiently Effective*	*Sufficiently Effective*
NASATLX Global		x		EEG Alpha Left Mean	x		
NASATLX Mental Demand	x			EEG Alpha Right Mean	x		
NASATLX Physical Demand		x		EEG Beta Left Mean	x		
NASATLX Temporal Demand		x		EEG Beta Right Mean	x		
NASATLX Effort	x			EEG Theta Left Mean	x		
NASATLX Frustration			x	EEG Theta Right Mean	x		
NASATLX Performance	x			TCD		x	
MRQ Auditory Emotional	x			TCD Left Mean		x	
MRQ Auditory Linguistic	x			TCD Right Mean	x		
MRQ Manual	x			fNIR			x
MRQ Short Term Memory	x			fNIR Left Mean		x	
MRQ Spatial Attentive		x		fNIR Right Mean		x	
MRQ Spatial Concentrative		x		ECG HR		x	
MRQ Spatial Categorical	x			ECG HRV	x		
MRQ Spatial Emergent	x			ECG IBI		x	
MRQ Spatial Positional		x					
MRQ Spatial Quantitative		x					
MRQ Visual Lexical	x						
MRQ Visual Phonetic	x						
MRQ Visual Temporal		x					
MRQ Vocal Process		x					
ISA	x						

Measure is significant at the 0.05 level

Measure is sufficiently effective at $\eta_2 \geq .138$

DISCUSSION

The analyses produced a number of interesting findings for interpreting in terms of the sufficiently effective standard for use on the three tasks types investigated within the present experiment, which occurred in the context of the nuclear domain. The approach for the discussion is similar to that of the results section, meaning a first look is at the typical human factors method most commonly presented in the literature to date. In other words, to develop an understanding of the impact of the results, the discussion below will begin with individual measure interpretation of ANOVAs in relation to meeting the sufficiently effective standard. Keeping line with the frequent style of reporting in the literature, an explanation for the correlations is provided. Going beyond the common reporting practices of workload findings, regression analyses are examined to inform the influence that workload measures had on the task type performance in which optimal performance with minimal errors are most critical in complex, high-risk environments. The conclusion provides an integrated interpretation of all findings followed by recommendations for future use of the measures and the sufficiency standard.

Effectiveness Checks

Numerous subjective and physiological measures were statistically significant at indicating workload changes between the three task types (checking, detection, and response implementation), yet few were sufficiently effective.

Subjective Measures

NASA-TLX.

Frustration was the only NASA-TLX subscale found to be sufficiently effective at indicating workload changes between the three task types, specifically frustration was highest in the detection task. At its foundation, the detection task stems from SDT. During the detection task, participants were required to monitor a gauge for five minutes and detect level changes by clicking on an acknowledge button located at the bottom of the display. Twelve random changes per minute occurred, totaling sixty gauge level changes per detection task. All four detection tasks occurred continuously. This required participants to monitor four different gauges consecutively for twenty minutes, totaling 240 gauge level changes. As a result, the detection tasks required participants to remain vigilant for a prolonged period. On the other hand, the checking and response implementation tasks, which exhibited lower frustration ratings, occurred rapidly and did not involve a prolonged period of vigilance. Past research supports this finding, showing that frustration reflects the primary workload component when performing vigilance task (Szalma, 2004; Warm, Dember, & Hancock, 1996).

In addition, Sawin and Scerbo (1995) found that NASA-TLX frustration levels increased in a vigilance task when participants received instruction that emphasized the importance of detecting signals by maintaining high levels of attentiveness (detection-emphasis) as opposed to relaxed instructions. In this experiment, when training for the detection task, the researcher provided participants with instructions that fall under the detection-emphasis category, further increasing each participant’s frustration level.

Physical demand was not sufficiently effective, but it was statistically significant. All three task types exhibited low amounts of physical demand, yet the detection task exhibited the highest rating. The only physical component to all three task types was mouse usage. The detection task required the highest amount of mouse usage. Although the response implementation task required more mouse usage than the checking task, they both required a minimal amount of mouse usage, thus they exhibited similar physical demand ratings.

Temporal Demand was not sufficiently effective, but it was statistically significant. All three task types exhibited high amounts of temporal demand, but the checking task exhibited the highest rating and the detection task exhibited the lowest rating. The pace in which participants performed each task drives this finding. As average time to complete the task block decreased, perceived temporal demand increased. The four checking tasks occurred over a 2-4 minute period; the four response implementation tasks occurred over a 3-5 minute period; and the four detection tasks occurred over a 20-25 minute period with each time period consisting of communication, navigation, and task execution. This finding indicates time pressure is a greater influence on the temporal demand scale of perceived workload as opposed to sustaining attention. That point informs cognitive trade-offs that might be occurring.

Mental demand, effort, and performance were not sufficiently effective or statistically significant at indicating workload changes between the three task types, but they all followed a trend similar to frustration and physical demand. For the detection task, mental demand and effort were highest and performance was lowest. Therefore, the trend supports hallmark patterns for vigilance tasks (Warm, Parasuraman, & Matthews, 2008), supporting that detection in the nuclear domain is a vigilance task. Both the checking and response implementation tasks displayed similar levels of mental demand, effort, and performance.

Taken as a whole the NASA-TLX findings indicate that the detection task elicited the highest level of workload, while the checking and response implementation tasks elicited similar levels of workload. The primary workload component for the detection task was frustration, which was also the only sufficiently effective measure at indicating workload changes between the three task types.

MRQ.

There were no sufficiently effective MRQ scales at indicating workload changes between the three task types. However, there were several statistically significant MRQ scales. The MRQ includes 17 items, but the developers of the questionnaire suggest removing items unrelated to the task (Boles & Adair, 2001). The researcher determined four items were unrelated to the present experiment and removed those four items. Of the 14 items that remained, six were statistically significant and eight were not statistically significant at indicating workload changes between the three task types. Based on the ratings of the MRQ scales, spatial concentrative, visual temporal, spatial quantitative, spatial attentive, spatial positional, and vocal process were statistically significant. Out of those six scales, four (spatial concentrative, visual temporal, spatial quantitative, spatial attentive) displayed similar results to the NASA-TLX. Ratings were highest in the detection task compared to the checking and response implementation tasks, which displayed comparable ratings.

Spatial concentrative processing was highest in the detection task. The design of the controls drives this finding. All tasks required a similar amount of spatial concentrative processing during the navigation component. However, once participants located the specific control needed to accomplish their task, the detection task required additional spatial concentrative processing. The detection task required participants to identify when a non-digital gauge reached a particular level. To complete this task, participants had to determine the numerical value of each dash by identifying the increments of the spaced dashes between gauge values. For example, one gauge contained the numbers 0, 500, 1000, 1500, 2000, 2500, and 3000, each with nine dashes in between. In this case, each dash was an increment of fifty. On the other hand, another gauge contained the numbers 0, 100, 200, 300, 400, 500, 600, and 700, each with four dashes in between. In this case, each dash was an increment of twenty. The differences in gauge design, coupled with the fact that the spacing between the dashes varied per gauge, attributed to a high amount of spatial concentrative processing during the detection task.

Spatial quantitative processing was highest in the detection task. The detection task required a high amount of spatial quantitative processing to determine each gauge level. The checking and response implementation task did not require identification of a numerical quantity.

Spatial attentive processing was highest in the detection task because of its vigilance component. The detection task required participants to focus their attention on a gauge for a prolonged period. On the other hand, once participants located the correct control during the checking and response implementation tasks, they could complete those tasks within a matter of seconds.

Visual temporal processing was highest in the detection task. The detection task was the only task that occurred for a predetermined amount of time. As a result, participants could have noticed that each detection task lasted the same amount of time and attempt to identify the task duration. In addition, participants could have attempted to identify any timing patterns between gauge level changes, in an effort to reduce their workload. Thus, it is well established that humans are not good predictors of time (Hancock, 1989), but the level of effort spent by the participants to attempt to accomplish this time determination appears to be an influencing factor to perceived workload. The checking and response implementation tasks did not occur for a predetermined amount of time and the steps within each task type occurred in rapid succession. Thus, participants could not identify the task duration or timing patterns.

Spatial positional processing was highest in the checking task. Per the definition of spatial positional processing, this finding indicates that the checking task required additional recognition of a precise location as differing from other locations compared to the detection and response implementation tasks. However, all three task types required the same amount of spatial processing. The difference between all three tasks is that the checking task solely consists of spatial positional processing, but spatial positional processing is only a single part of the detection and response implementation tasks. As a result, during the checking task, participants allocated all of their resources to spatial positional processing, whereas during the detection and response implementation tasks, participants allocated only part of their available resources to spatial positional processing.

Vocal processing was highest in the response implementation and checking tasks. All three task types required a similar amount of voice usage via three-way communication. However, participants executed the checking and response implementation tasks more quickly, producing a shorter break between the two three-way communication parts. As a result, during the checking and response implementation tasks, participants were communicating with the SRO at a faster pace. The detection task consisted of a five-minute period where participants monitored a control without communication with the SRO, thus they received a five-minute communication break between the two three-way communication parts. As a result, during the detection task type, participants were communicating with the SRO at a slower pace. In addition, the vocal processing finding indicates that the working memory component required for the communication reporting for the detection task type was limited or non-existent in its influence.

Of the 14 items used, the eight items that were not statistically significant were auditory emotional, auditory linguistic, manual process, short-term memory, spatial categorical, spatial emergent, visual lexical, and visual phonetic. Auditory emotional, auditory linguistic, and short-term memory were specifically added to identify the processing required during three-way communication and results showed that those were not affected by the task type. This finding indicates that the processing required to complete the three-way communication portion of each task was consistent.

Spatial categorical, spatial emergent, visual lexical and visual phonetic were specifically added to identify the processing needed to locate controls on each panel. Results indicate that the task type manipulation did not affect these processes. Therefore, the MRQ was sensitive at capturing the effects of task type manipulation on visual and spatial processing without interference from other factors, such as panel design, supporting the present experiment methodology.

Manual process was specifically added to identify the physical arm, hand, and finger movement between the three task types. Contrary to NASA-TLX physical demand, results indicate that MRQ manual process was not sensitive to the task type manipulation. The NASA-TLX asked participants how physically demanding the task was. The MRQ asked participants to rate the task on the extent to which they used each process. The NASA-TLX physical demand findings indicate that the tasks required different levels of physical demand, yet all three tasks required a low amount of physical demand. Because the task type required low amounts of physical demand, there was no difference in the amount of manual process required to complete each task.

Taken as a whole, the MRQ findings indicate that the detection task elicited the highest amount of processing for most of the statistically significant measures. However, out of the 14 measures used, none were sufficiently effective at indicating workload changes between the three task types.

ISA.

The ISA was not sufficiently effective or statistically significant at capturing changes in workload between the three task types. This finding is attributed to two factors in conjunction: the online nature of the rating and its lack of diagnosticity. Prior research has suggested that the ISA is a sensitive measure of workload (Abich, 2013; Tattersall & Ford, 1996), but these research efforts have investigated the ISA's sensitivity to variations in task difficulty. This research effort investigated the ISA's sensitivity to variations across task types. This difference highlights a problem with the online nature of the ISA and its lack of diagnosticity. When using the ISA to investigate variations in task difficulty, researchers expose participants to the same task while manipulating task difficulty. This process exposes participants to the task for a prolonged period. This prolonged task exposure provides a reference for enabling participants to make an educated online assessment of their workload and does not require sub questions that make other subjective workload measures diagnostic. However, when using the ISA to investigate differences across task types, researchers expose participants to different task types for smaller periods. In these instances, participants are not equipped with enough task exposure. Without prolonged task exposure and a lack of sub questions to act as a memory trigger, participants do not have enough information for comparison to make an educated online assessment of their workload. Therefore, the ISA is not effective at indicating changes in workload across task types.

Performance Measures.

Instruction performance supports many of the subjective and physiological workload findings. Percent correct was lowest during the detection task, while percent correct on both the checking and response implementation tasks was high. The difference in percent correct between the checking and response implementation tasks was negligible. Instruction clarification and request for repeat instructions followed the same trend. Both were highest in the detection task, while the difference between the checking and response implementation task types was negligible.

All three task types required the same amount of communication steps, but the instruction from the SRO to the RO varied within those steps. These differences in instruction led to a performance difference. The instructions during the checking and response implementation tasks were similar. For example, during the checking task the SRO would instruct the RO to "verify valve PCV-444B is shut." During the response implementation task the SRO would instruct the RO to "shut valve 1CS-235B." On the other hand, the detection task consisted of a longer instruction. During the detection task the SRO would instruct the RO to "verify gauge TI-430 SB and report when less than 400 PSIG."

Overall performance, which was a combination of instruction performance and execution performance, showed a performance workload dissociation for the checking and response implementation tasks. Overall performance was similar across all three task types. The unexpected mediocre overall performance on the checking and response implementation tasks stems from execution performance. Participants executed the checking task correctly 61% of the time. Although the checking task required fewer steps than the detection and response implementation tasks, participants often skipped the physical step of clicking on the correct control. Participants executed the response implementation task correctly 60% of the time. Although the response implementation task only required two steps, participants often skipped the physical step of clicking the control before opening or shutting the valve. All three tasks required a physical action using the mouse, but participants could not complete the detection task without clicking the control via the mouse. The checking task required participants to click the correct control once they located it, but participants could complete the task without clicking on the correct control. The response implementation task required participants to click on the correct control before they opened or shut the valve, but participants could manipulate the valve without the initial click on the control. Participants often forgot to click on the control for identification, but could still complete both the checking and response implementation tasks without clicking on the control. The identification click is important in an NPP MCR because operators point to the controls to allow for back-up behavior from the SRO. Therefore, the operational relevance makes this step critical for determining and informing overall performance.

Physiological Measures

EEG.

The EEG measures were not sufficiently effective or statistically significant at capturing changes in frequency band for each hemisphere, lobe, and channel between the three task types.

The task demands did not require a change in brain electrical activity that was detectable by this measure.

Two factors attributed to this finding. First, the cognitive and physical loads between all three task types were analogous for the EEG to detect differences. Second, all three task types did not invoke a cognitive and physical load high enough to be detected by EEG. NASA-TLX supports this notion by showing that global workload rating for all three task types was below 40.

EEG might be a valuable measure that detects differences between low, medium, and high workload (Abich, 2013; Brookings, Wilson, Swain, 1996; Brouwer et al., 2012; Putze, Jarvis, Schultz, 2010) or medium and high workload (Brouwer et al, 2012; Wilson, 2002), but ineffective at detecting differences between tasks with different variations of low workload.

TCD.

The TCD measures were not a sufficiently effective measure at indicating workload changes between the three task types, but left hemisphere CBFV was statistically significant. When comparing all three task types, left hemisphere CBFV was highest for the checking task. In fact, left hemisphere CBFV increased from baseline during the checking task but decreased from baseline during the detection and response implementation tasks. However, when further examining left hemisphere CBFV data, standard error presented an issue (see figure 13). The task type manipulation did cause changes in left hemisphere CBFV that was detectable by the TCD, but the variance in the data set reduces the merit of the results. Thus, TCD should be considered ineffective at indicating workload changes between the three task types.

fNIR.

fNIR measures were sufficiently effective at indicating workload changes between the three task types. When comparing all three task types, blood oxygenation difference from baseline decreased for all three tasks, but showed the least change from baseline for the detection task. The prefrontal cortex is associated with planning cognitive behavior and decision-making. As supported by the MRQ, the detection task required more planning and decision-making compared to the checking and response implementation tasks. During the detection task, participants first had to locate the gauge of interest. Once participants located the gauge, they had to determine the current level of the gauge and where the gauge level needed to move to before the task was complete. Once these three steps were complete, participants began detecting gauge level changes for five minutes. On the other hand, the checking task required fewer steps. During the checking task, participants had to locate the control of interest and then determine its currents state. The response implementation task required one more action than the checking task, but still had fewer actions than the detection task. The extra action required participants to open or shut a control via the mouse, which is a fine motor task. Therefore, the difference seen in the fNIR responses is reflective of the task complexity demonstrated by the number of actions required.

fNIR affords good spatial localization compared to EEG, which is why fNIR was sufficiently effective at indicating workload changes in the prefrontal cortex, but frontal lobe EEG was insufficiently effective (Gratton & Fabiani, 2008).

ECG.

HR, HRV, and IBI were not sufficiently effective measures at indicating workload changes between the three task types, but HR and IBI were statistically significant.

When comparing all three task types, HR difference from baseline increased for all three tasks, but was greatest in the checking task and least in the detection task. IBI difference from baseline increased for all three tasks, but was greatest in the checking and response implementation tasks and least in the detection task. Logically, increasing HR would lead to a decrease in the time between heartbeats. These findings are similar to that of the NASA-TLX temporal demand and MRQ vocal process rating. Participants felt high temporal demand and used a high amount of vocal processing during the checking task because they were communicating with the SRO at a fast pace. The faster paced communication is reflected by immediate arousal and an increased heart rate.

Correlations

The three correlation analyses indicate that the relationship between subjective and objective measures is task dependent. Results indicated negative relationships between subjective and objective measures for the checking and detection tasks and positive relationships for the response implementation task. This finding is evidence that subjective and objective measures are not measuring the same construct.

There were no positive or negative correlations between a subjective and objective measure consistently present across all three tasks types. However, certain correlations were present across pairs of tasks. Left hemisphere CBFV was positively correlated with MRQ manual process for the detection and response implementation tasks. Left hemisphere CBFV was positively correlated with MRQ vocal process for the checking and response implementation tasks. Right hemisphere blood oxygenation was negatively correlated with MRQ spatial quantitative processing for the detection and response implementation tasks. Interestingly, ISA was not correlated with anything, providing further evidence that it does not measure workload for these three task types.

Regressions

The findings from the correlations led to the thinking that these measures might independently account for different amounts of variance in explaining performance, which provided support for including both subjective and objective measures in a regression model. Results of the three regressions, calculated for each task type overall performance, indicate that task difficulty influences which variables are predictors of performance. Subjective measures entered at Step 1 were not significant predictors of performance on the response implementation task, and when physiological measures were entered at Step 2, the model was still not significant. Subjective measures entered at Step 1 were significant predictors of performance on the checking task, but when the physiological measures were entered at Step 2 the model was not significant. For the detection task, subjective measures entered at Step 1 were significant predictors of performance, and when physiological measures were entered at Step 2 the model was significant and R^2 increased. These results indicate that as task complexity increases so does the ability of physiological measures to predict performance.

Of the two sufficiently effective measures that indicated workload changes between the three task types, only the NASA-TLX frustration contributed significantly to the model for predicting performance as indicated by the beta weights. NASA-TLX frustration contributed significantly to the model for predicting performance during both Step 1 and 2 of the detection task. fNIR was not a significant predictor of performance on any task. These findings indicate that while the fNIR was sufficiently effective, the NASA-TLX actually yields the greatest utility, keeping in mind that predicting performance is the ultimate goal for understanding and utilizing workload measures in the present domain. Predicting performance is the goal strived for because safety of the utmost importance and successful task completion through optimal performance is the means to maintain public safety and health.

Of the statistically significant but not sufficiently effective measures, MRQ vocal process and ECG HR contributed significantly to the model for predicting performance. MRQ vocal process contributed significantly to the model for Step 2 of the checking task. ECG contributed significantly to the model for Step 2 of the detection task. Those findings are important to consider because even though those measures were not meet the standard to be sufficiently effective, they still appear to be useful in predicting performance. Interestingly, two of the measures that were not statistically significant at indicating workload changes between the three task types contributed significantly to the model for predicting performance. EEG Theta P4significantly contributed to the model of performance during Step 2 of the checking task. The ISA P4 significantly contributed to the model of performance during Step 1 and 2 of the detection task. These findings indicate that despite the fact EEG and ISA were not effective at indicating changes in workload between the three task types, both have utility as measures that predict performance.

Taking all of the findings into account from this section, it seems as though predicting performance needs to be the priority and to do so should not be limited to strictly theoretically derived variables or mathematically derived variables, but perhaps a combination of both. In other words, researchers need to determine the objective for their research prior to analysis execution to guide the variable selection. For example, deciding on whether the question of importance is understanding workload measures for task development in experimentation (what measures to use) versus predicting performance for system design and task allocation.

Conclusion

The purpose for the present experiment was to determine if certain workload measures are sufficiently effective across domains by taking the findings from one domain (military) and testing whether those results hold true in a different domain (nuclear). As detailed earlier in the discussion section, only two measures were sufficiently effective at indicating workload changes between the three task types in the nuclear domain, but many measures were statistically significant. The effect size in the sufficiency standard developed for this research effort was based on the premise that all three task types have varying taskloads. This premise was based on a task analysis conducted by Reinerman-Jones et al. (2013), which revealed the different components of all three task types. Because all three task types are composed of different components, each requires different processing and potentially have varying taskloads. However, both subjective and objective results indicated that the checking and response implementation tasks elicited similar levels of workload. This finding resulted in lower than expected effect sizes, thus deeming many measures not sufficiently effective.

In the military domain, Abich (2013) found that the ISA and HRV were sensitive to task demands across three different task types, thus determining that several subjective and objective measures were not efficient at indicating workload changes between task types within the same domain. The results of the present research effort combined with the results from Abich (2013) highlight an alarming problem: the inability of a consistent subjective and physiological measure to indicate changes in workload across tasks (Abich, 2013) and across domains. This research effort concluded that both the ISA and HRV, which Abich (2013) recommended, were not sufficiently effective or even statistically significant at indicating workload changes. Consequently, based on multiple research efforts, no single measure was effective at indicating workload changes across the military (ISR operations) and nuclear (NPP MCR operations) domains.

An additional finding based on the results of this research effort and Abich (2013) is the inconsistency of subjective and physiological measures to predict performance for theoretically similar tasks across domains. Both this research effort and Abich (2013) included tasks founded on SDT. When predicting performance for his threat detection task, Abich (2013) regressed performance for a single level of task demand on workload measures for all three levels of demand for that task. This resulted in nine regressions. Based on the results of the present research effort, the detection task falls under the low/medium workload range for the given event rate manipulation. Consequently, when comparing results it is appropriate to use the results of Abich's (2013) regressions for the low and medium task demand on workload measures for the threat detection task. Abich (2013) found that both low and medium subjective and physiological measures did not predict performance for the low and medium demand threat detection task. On the other hand, this research effort found that subjective and physiological measures predict performance during the detection task and that the NASA-TLX frustration and ISA were significantly contributed to the prediction model of performance during both steps.

The take away from Abich's (2013) research was that both subjective and objective measures are consistently inconsistent at measuring workload and predicting performance across different tasks within the same domain. This research effort found that the same subjective and objective measures are consistently inconsistent at measuring workload across domains. This suggests that workload is a multidimensional construct with multifaceted factors that affect the construct. These factors are task and domain specific. A single measure is unable to capture the complex construct of workload across different tasks within the same domain or across domains. Thus, workload might be the operator's perceived evaluation to the experience imposed by the task demands and physiological response to the task components themselves. If the end goal of using workload measures is to advance them to real world spaces, the approach of modern day research is inadequate. This research effort highlights the importance of proper methodology. As researchers, we have to identify the appropriate workload measure for all tasks regardless of the domain by investigating the effectiveness of each measure. The findings of the present study suggest that responsible science include evaluating workload measures before use, not relying on prior research or theory. In other words, results indicate that it is only acceptable to use a measure based on prior findings if research has tested that measure on the exact task and manipulations within that specific domain.

Recommendations

The best recommendations available for workload measures in the nuclear domain are resultant of this research effort, which was the first to guide future research in the NPP domain by identifying the workload measures that are sufficiently effective at indicating changes in workload in common NPP MCR tasks. Recommendations drawn for the present experiment suggest that when identifying workload changes between the three primary task types performed in NPP MCR, the NASA-TLX is the best subjective measure of workload. Although there were no sufficiently effective subjective workload measures, the NASA-TLX was the only measure with a sufficiently effective subscale. In addition, half of the NASA-TLX subscales were statistically significant at identifying workload differences between the three task types. Less than half of the MRQ subscales were statistically significant and the ISA was not statistically significant.

Recommendations drawn for the present experiment suggest that when identifying workload changes between the three primary task types performed in NPP MCR, fNIR is the ideal physiological measure of workload, as it was the only sufficiently effective measure. TCD and ECG were statistically significant at identifying workload changes between the three task types, but effect sizes were low, variance was high, and results were inconsistent with all other findings. fNIR also proved to be the best overall measure of workload at identifying differences between task types.

Future research efforts examining the primary task performed by reactor operators in NPP MCRs should consider revising the sufficiency standard, as it might be too stringent. Modifying the required effect size by utilizing the effect sizes provided in this research effort would create a more suitable sufficiency standard. This is not to say that the sufficiency standard created for this research effort is inadequate. The sufficiency standard is better suited at measuring the effectiveness of workload measures for tasks with distinct difference in task demand, such as the research conducted by Abich (2013).

Future research should also examine the effectiveness of the subjective and objective measures utilized in this research effort and that of Abich (2013) at indicating workload changes in other domains. As highlighted in the introduction, domains such as medicine, aviation, military, and nuclear can be characterized as similar structures involving an operator or a team of operators/personnel. Research should expand to identify the appropriate workload measures to utilize for primary tasks performed by operators/personnel in the medical and aviation domains.

In addition, it should be noted that this research and that by Abich (2013) examined primarily rule-based tasks and as such, workload measures appear to simply reflect task requirements that might more easily be derived from a task analysis. Therefore, future research needs to investigate these measures for assessing workload elicited by tasks requiring a knowledge-base. Abich and this research effort clearly highlight the need to question and investigate the goodness of fit of measures before inclusion in experimentation and making concrete determinations based on resultant data.

APPENDIX A:
SUBJECTIVE AND OBJECTIVE CORRELATIONS

Table A1

Subjective and Objective Correlations for the Checking Task Type

	NASATLX Mental Demand	*NASATLX Physical Demand*	*NASATLX Temporal Demand*	*NASATLX Effort*	*NASATLX Frustration*	*NASATLX Performance*	*MRQ Auditory Emotional*	*MRQ Auditory Linguistic*	*MRQ Manual*	*MRQ Short Term Memory*	*MRQ Spatial Attentive*
NASATLX Mental Demand	1										
NASATLX Physical Demand	.267*	1									
NASATLX Temporal Demand	.631**	.390**	1								
NASATLX Effort	.770**	.448**	.644**	1							
NASATLX Frustration	.396**	.061	.441**	.376**	1						
NASATLX Performance	.169	.002	.300**	.212	.434**	1					
MRQ Auditory Emotional	.242*	.220*	.276*	.254*	.049	.232*	1				
MRQ Auditory Linguistic	.168	.155	.109	.178	-.257*	-.143	.058	1			
MRQ Manual	.294**	.452**	.227*	.379**	.085	.068	.260*	.333**	1		
MRQ Short Term Memory	.383**	.079	.338**	.425**	.171	.022	.058	.478**	.441**	1	
MRQ Spatial Attentive	.520**	.243*	.300**	.476**	.296**	.102	.114	.386**	.463**	.622**	1
MRQ Spatial Concentrative	.392**	.336**	.299**	.349**	.299**	.106	.356**	.122	.334**	.315**	.459**
MRQ Spatial Categorical	.303**	.170	.304**	.222*	.215	.209	.229*	.154	.300**	.169	.469**
MRQ Spatial Emergent	.213	.249*	.182	.332**	.169	-.024	.187	.232*	.450**	.400**	.531**
MRQ Spatial Positional	.302**	.208	.314**	.333**	.152	.105	.165	.462**	.506**	.509**	.702**
MRQ Spatial Quantitative	.182	.142	.238*	.155	.035	.085	.351**	.272*	.233*	.281*	.303**
MRQ Visual Lexical	.293**	.228*	.209	.357**	-.088	-.009	.116	.568**	.449**	.553**	.469**
MRQ Visual Phonetic	.152	.357**	.304**	.157	.060	.183	.423**	.360**	.256*	.215	.395**
MRQ Visual Temporal	.296**	.262*	.382**	.252*	.287**	.221*	.562**	.157	.308**	.207	.262*
MRQ Vocal Process	.106	.140	.143	.162	-.218	-.059	.154	.477**	.313**	.365**	.293**
ISA	.256*	-.005	.303**	.395**	.448**	.306**	-.014	-.009	.094	.244*	.151

Variables	*NASATLX Mental Demand*	*NASATLX Physical Demand*	*NASATLX Temporal Demand*	*NASATLX Effort*	*NASATLX Frustration*	*NASATLX Performance*	*MRQ Auditory Emotional*	*MRQ Auditory Linguistic*	*MRQ Manual*	*MRQ Short Term Memory*	*MRQ Spatial Attentive*
EEG Alpha Left Mean	.171	-.038	.134	.138	.119	.110	.036	.119	.124	.028	.063
EEG Alpha Right Mean	-.286*	-.052	-.242*	-.160	-.091	-.178	-.160	-.131	-.214	-.124	-.163
EEG Beta Left Mean	.171	-.042	.132	.136	.119	.111	.037	.119	.122	.028	.063
EEG Beta Right Mean	- .286*	-.056	-.245*	-.162	-.090	-.177	-.159	-.131	-.214	-.124	-.163
EEG Theta Left Mean	.171	-.040	.132	.135	.117	.109	.038	.119	.122	.026	.062
EEG Theta Right Mean	-.286*	-.058	-.246*	-.163	-.090	-.177	-.158	-.132	-.216	-.127	-.164
TCD Left Mean	-.027	-.140	-.064	.028	-.113	-.002	.073	.052	.121	.080	.049
TCD Right Mean	-.052	-.012	-.001	-.148	.232*	.068	.142	-.265*	-.028	-.080	-.026
fNIR Left Mean	.122	-.042	-.042	.075	.092	.064	.174	-.050	.149	-.084	.023
fNIR Right Mean	.028	-.050	-.025	.038	.000	-.072	.155	-.006	.190	.026	-.005
ECG HR	-.173	-.034	.011	-.084	-.167	-.067	-.075	-.143	-.284*	-.304**	-.260*
ECG HRV	-.188	.037	.006	-.170	-.196	.053	.098	.193	.132	.036	-.167
ECG IBI	.135	-.072	-.041	.085	.155	-.033	-.023	.086	.161	.225	.222

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Variables	*MRQ Spatial Concentrative*	*MRQ Spatial Categorical*	*MRQ Spatial Emergent*	*MRQ Spatial Positional*	*MRQ Spatial Quantitative*	*MRQ Visual Lexical*	*MRQ Visual Phonetic*	*MRQ Visual Temporal*	*MRQ Vocal Process*	*ISA*	*EEG Alpha Left Mean*
NASATLX Mental Demand											
NASATLX Physical Demand											
NASATLX Temporal Demand											
NASATLX Effort											
NASATLX Frustration											
NASATLX Performance											
MRQ Auditory Emotional											
MRQ Auditory Linguistic											
MRQ Manual											
MRQ Short Term Memory											
MRQ Spatial Attentive											
MRQ Spatial Concentrative	1										
MRQ Spatial Categorical	.423**	1									
MRQ Spatial Emergent	.431**	.285**	1								
MRQ Spatial Positional	.312**	.354**	.640**	1							
MRQ Spatial Quantitative	.515**	.462**	.332**	.446**	1						
MRQ Visual Lexical	.240*	.229*	.288**	.489**	.362**	1					
MRQ Visual Phonetic	.433**	.432**	.294**	.460**	.584**	.415**	1				
MRQ Visual Temporal	.509**	.413**	.235*	.263*	.476**	.216	.548**	1			
MRQ Vocal Process	.049	.028	.291**	.401**	.165	.372**	.217	.021	1		
ISA	.208	.197	.094	.036	.011	.039	.091	.282*	-.058	1	

Variables	*MRQ Spatial Concentrative*	*MRQ Spatial Categorical*	*MRQ Spatial Emergent*	*MRQ Spatial Positional*	*MRQ Spatial Quantitative*	*MRQ Visual Lexical*	*MRQ Visual Phonetic*	*MRQ Visual Temporal*	*MRQ Vocal Process*	*ISA*	*EEG Alpha Left Mean*
EEG Alpha Left Mean	.040	.179	.197	.059	.027	-.028	-.139	.100	.178	.109	1
EEG Alpha Right Mean	-.079	-.202	-.083	-.152	-.142	-.164	-.156	-.259*	-.003	-.005	-.050
EEG Beta Left Mean	.038	.181	.196	.058	.027	-.028	-.138	.100	.177	.108	1
EEG Beta Right Mean	-.081	-.199	-.084	-.152	-.143	-.164	-.156	-.259*	-.002	-.004	-.044
EEG Theta Left Mean	.039	.180	.195	.058	.026	-.028	-.139	.100	.178	.107	1
EEG Theta Right Mean	-.082	-.198	-.085	-.154	-.143	-.165	-.157	-.259*	-.003	-.005	-.042
TCD Left Mean	-.022	-.147	.071	.206	-.022	.062	-.074	-.101	.243*	-.083	.092
TCD Right Mean	.194	.173	.077	-.068	.003	-.168	-.014	.165	-.110	-.030	.175
fNIR Left Mean	.186	.005	.090	.009	-.054	.091	-.069	.085	-.039	.012	.130
fNIR Right Mean	.031	.003	.157	.010	-.118	.051	-.096	.074	-.012	-.008	.023
ECG HR	-.155	-.115	-.259*	-.188	-.117	-.107	.072	-.011	-.104	-.063	-.182
ECG HRV	-.073	-.188	.027	.006	.025	.075	.014	.159	.098	-.062	.182
ECG IBI	.080	.063	.187	.104	.070	.054	-.024	-.037	-.016	.074	.162

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Variables	*EEG Alpha Right Mean*	*EEG Beta Left Mean*	*EEG Beta Right Mean*	*EEG Theta Left Mean*	*EEG Theta Right Mean*	*TCD Left Mean*	*TCD Right Mean*	*fNIR Left Mean*	*fNIR Right Mean*	*ECG HR*	*ECG HRV*	*ECG IBI*
NASATLX Mental Demand												
NASATLX Physical Demand												
NASATLX Temporal Demand												
NASATLX Effort												
NASATLX Frustration												
NASATLX Performance												
MRQ Auditory Emotional												
MRQ Auditory Linguistic												
MRQ Manual												
MRQ Short Term Memory												
MRQ Spatial Attentive												
MRQ Spatial Concentrative												
MRQ Spatial Categorical												
MRQ Spatial Emergent												
MRQ Spatial Positional												
MRQ Spatial Quantitative												
MRQ Visual Lexical												
MRQ Visual Phonetic												
MRQ Visual Temporal												
MRQ Vocal Process												
ISA												

Variables	*EEG Alpha Right Mean*	*EEG Beta Left Mean*	*EEG Beta Right Mean*	*EEG Theta Left Mean*	*EEG Theta Right Mean*	*TCD Left Mean*	*TCD Right Mean*	*fNIR Left Mean*	*fNIR Right Mean*	*ECG HR*	*ECG HRV*	*ECG IBI*
EEG Alpha Left Mean												
EEG Alpha Right Mean	1											
EEG Beta Left Mean	-.050	1										
EEG Beta Right Mean	1	-.044	1									
EEG Theta Left Mean	-.051	1	-.045	1								
EEG Theta Right Mean	1	-.041	1	-.043	1							
TCD Left Mean	.076	.092	.077	.090	.076	1						
TCD Right Mean	.057	.175	.059	.175	.059	.283*	1					
fNIR Left Mean	-.144	.128	-.145	.129	-.145	.202	.240*	1				
fNIR Right Mean	-.239*	.023	-.239*	.023	-.240*	.232*	.209	.658**	1			
ECG HR	.204	-.184	.202	-.182	.203	-.113	-.195	-.056	-.117	1		
ECG HRV	-.119	.181	-.120	.181	-.121	.073	-.078	-.045	.074	-.239*	1	
ECG IBI	-.218	.164	-.216	.161	-.217	.169	.248*	.077	.160	-.644**	.016	1

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Table A2

Subjective and Objective Correlations for the Detection Task Type

Variables	*NASATLX Mental Demand*	*NASATLX Physical Demand*	*NASATLX Temporal Demand*	*NASATLX Effort*	*NASATLX Frustration*	*NASATLX Performance*	*MRQ Auditory Emotional*	*MRQ Auditory Linguistic*	*MRQ Manual*	*MRQ Short Term Memory*	*MRQ Spatial Attentive*
NASATLX Mental Demand	1										
NASATLX Physical Demand	.411**	1									
NASATLX Temporal Demand	.551**	.463**	1								
NASATLX Effort	.759**	.421**	.592**	1							
NASATLX Frustration	.284*	.230*	.299**	.343**	1						
NASATLX Performance	.155	-.048	.110	.192	.005	1					
MRQ Auditory Emotional	.110	.314**	.218	.081	.221*	.212	1				
MRQ Auditory Linguistic	.138	.208	.129	.244*	-.159	-.080	.164	1			
MRQ Manual	.223*	.644**	.390**	.426**	.154	-.049	.097	.321**	1		
MRQ Short Term Memory	.223*	.120	.112	.246*	.038	-.140	.045	.564**	.240*	1	
MRQ Spatial Attentive	.106	.181	.181	.171	.058	-.163	-.009	.216	.151	.381**	1
MRQ Spatial Concentrative	.259*	.264*	.325**	.316**	-.029	.021	.265*	.255*	.091	.343**	.506**
MRQ Spatial Categorical	-.041	.092	.135	.009	.024	-.094	.145	.026	.084	.097	.377**
MRQ Spatial Emergent	.148	.148	.260*	.059	-.123	-.162	.295**	.350**	.207	.369**	.326**
MRQ Spatial Positional	.003	.129	.147	.075	-.121	-.044	.086	.124	.130	.173	.675**
MRQ Spatial Quantitative	.123	.179	.030	.103	-.107	-.120	.089	.171	-.053	.142	.507**
MRQ Visual Lexical	.076	.337**	.252*	.170	-.001	-.222*	-.032	.330**	.484**	.292**	.403**
MRQ Visual Phonetic	.229*	.302**	.283*	.270*	-.070	.094	.205	.279*	.286**	.312**	.379**
MRQ Visual Temporal	.296**	.322**	.197	.296**	.073	.208	.391**	.163	.153	.059	.147
MRQ Vocal Process	.110	.230*	.037	.200	-.147	-.149	.063	.638**	.331**	.423**	.177
ISA	.295**	.206	.394**	.294**	.192	.100	.061	.071	.138	.234*	.001

Variables	*MRQ Spatial Concentrative*	*MRQ Spatial Categorical*	*MRQ Spatial Emergent*	*MRQ Spatial Positional*	*MRQ Spatial Quantitative*	*MRQ Visual Lexical*	*MRQ Visual Phonetic*	*MRQ Visual Temporal*	*MRQ Vocal Process*	*ISA*	*EEG Alpha Left Mean*
NASATLX Mental Demand											
NASATLX Physical Demand											
NASATLX Temporal Demand											
NASATLX Effort											
NASATLX Frustration											
NASATLX Performance											
MRQ Auditory Emotional											
MRQ Auditory Linguistic											
MRQ Manual											
MRQ Short Term Memory											
MRQ Spatial Attentive											
MRQ Spatial Concentrative	1										
MRQ Spatial Categorical	.342**	1									
MRQ Spatial Emergent	.438**	.287**	1								
MRQ Spatial Positional	.407**	.499**	.467**	1							
MRQ Spatial Quantitative	.415**	.382**	.200	.472**	1						
MRQ Visual Lexical	.251*	.354**	.259*	.361**	.284*	1					
MRQ Visual Phonetic	.358**	.218	.233*	.413**	.374**	.446**	1				
MRQ Visual Temporal	.386**	.128	.110	.166	.398**	.138	.451**	1			
MRQ Vocal Process	.065	.071	.246*	.150	.260*	.360**	.319**	.275*	1		
ISA	.021	-.185	-.030	-.115	-.044	.066	.126	.056	.082	1	

Variables	EEG Alpha Right Mean	EEG Beta Left Mean	EEG Beta Right Mean	EEG Theta Left Mean	EEG Theta Right Mean	TCD Left Mean	TCD Right Mean	fNIR Left Mean	fNIR Right Mean	ECG HR	ECG HRV	ECG IBI
NASATLX Mental Demand												
NASATLX Physical Demand												
NASATLX Temporal Demand												
NASATLX Effort												
NASATLX Frustration												
NASATLX Performance												
MRQ Auditory Emotional												
MRQ Auditory Linguistic												
MRQ Manual												
MRQ Short Term Memory												
MRQ Spatial Attentive												
MRQ Spatial Concentrative												
MRQ Spatial Categorical												
MRQ Spatial Emergent												
MRQ Spatial Positional												
MRQ Spatial Quantitative												
MRQ Visual Lexical												
MRQ Visual Phonetic												
MRQ Visual Temporal												
MRQ Vocal Process												
ISA												

Variables	*NASATLX Mental Demand*	*NASATLX Physical Demand*	*NASATLX Temporal Demand*	*NASATLX Effort*	*NASATLX Frustration*	*NASATLX Performance*	*MRQ Auditory Emotional*	*MRQ Auditory Linguistic*	*MRQ Manual*	*MRQ Short Term Memory*	*MRQ Spatial Attentive*
EEG Alpha Left Mean	-.103	.027	-.154	-.132	.115	.084	.183	-.159	.046	-.116	-.107
EEG Alpha Right Mean	.065	.000	-.012	.028	.093	.049	-.019	-.066	-.035	-.115	.094
EEG Beta Left Mean	-.104	.027	-.155	-.133	.116	.082	.182	-.158	.047	-.115	-.107
EEG Beta Right Mean	.064	.000	-.013	.028	.095	.050	-.019	-.062	-.032	-.112	.095
EEG Theta Left Mean	-.100	.028	-.148	-.129	.110	.089	.187	-.158	.045	-.118	-.107
EEG Theta Right Mean	.065	-.005	-.012	.026	.093	.050	-.015	-.072	-.041	-.118	.094
TCD Left Mean	.048	.041	.059	.170	-.001	.008	-.120	.031	.293**	-.005	.029
TCD Right Mean	-.053	-.160	-.223	-.153	-.023	-.062	-.056	-.120	-.326**	-.093	-.074
fNIR Left Mean	.053	.085	.159	.080	.022	.040	-.020	.020	.264*	-.055	-.127
fNIR Right Mean	.135	.026	.077	.113	.126	-.059	-.137	-.067	.239*	-.001	-.215
ECG HR	.036	.102	.264*	.039	-.001	-.148	-.052	.019	.063	.042	.044
ECG HRV	-.002	-.163	.013	-.021	-.126	.078	-.001	.083	-.034	-.017	.020
ECG IBI	-.335**	-.131	-.211	-.329**	-.009	.029	-.013	-.383**	-.130	-.258*	-.015

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Variables	*MRQ Spatial Concentrative*	*MRQ Spatial Categorical*	*MRQ Spatial Emergent*	*MRQ Spatial Positional*	*MRQ Spatial Quantitative*	*MRQ Visual Lexical*	*MRQ Visual Phonetic*	*MRQ Visual Temporal*	*MRQ Vocal Process*	*ISA*	*EEG Alpha Left Mean*
EEG Alpha Left Mean	-.169	-.155	-.176	-.149	-.215	-.088	-.063	.158	.000	.053	1
EEG Alpha Right Mean	.103	.016	.002	.003	-.012	-.003	-.161	-.052	-.026	.020	.440**
EEG Beta Left Mean	-.170	-.155	-.176	-.149	-.214	-.085	-.062	.157	.001	.052	1
EEG Beta Right Mean	.103	.020	.003	.003	-.014	-.002	-.160	-.050	-.026	.020	.438**
EEG Theta Left Mean	-.167	-.153	-.175	-.145	-.222*	-.099	-.064	.161	-.001	.055	.999**
EEG Theta Right Mean	.105	.014	.002	.005	-.011	-.007	-.154	-.046	-.027	.014	.444**
TCD Left Mean	-.097	.029	.062	.151	.069	.196	.044	.056	.084	.050	-.152
TCD Right Mean	-.056	.123	.088	.042	.048	-.168	-.118	-.036	-.096	-.285*	.058
fNIR Left Mean	-.077	-.095	.101	-.007	-.184	.164	.036	-.131	-.018	.076	.055
fNIR Right Mean	-.136	-.137	-.016	-.231*	-.242*	-.045	-.217	-.120	-.070	.018	.108
ECG HR	.065	-.033	.029	-.009	.010	.004	.111	-.096	-.053	.142	-.061
ECG HRV	.110	-.151	.072	.056	.123	-.011	.127	.181	.093	.256*	.058
ECG IBI	-.184	.130	-.139	-.004	-.150	.013	-.226	-.114	-.194	-.167	.194

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Variables	*EEG Alpha Right Mean*	*EEG Beta Left Mean*	*EEG Beta Right Mean*	*EEG Theta Left Mean*	*EEG Theta Right Mean*	*TCD Left Mean*	*TCD Right Mean*	*fNIR Left Mean*	*fNIR Right Mean*	*ECG HR*	*ECG HRV*	*ECG IBI*
EEG Alpha Left Mean												
EEG Alpha Right Mean	1											
EEG Beta Left Mean	.440**	1										
EEG Beta Right Mean	1	.438**	1									
EEG Theta Left Mean	.440**	.999**	.438**	1								
EEG Theta Right Mean	.999**	.445**	.998**	.444**	1							
TCD Left Mean	-.223	-.153	-.222	-.154	-.224*	1						
TCD Right Mean	-.061	.058	-.060	.057	-.057	.204	1					
fNIR Left Mean	.034	.054	.033	.055	.033	-.019	-.173	1				
fNIR Right Mean	-.046	.107	-.047	.107	-.048	.073	.142	.553**	1			
ECG HR	-.093	-.061	-.092	-.063	-.095	.042	-.013	-.207	-.178	1		
ECG HRV	.119	.056	.117	.061	.119	.134	-.023	.186	.097	-.229*	1	
ECG IBI	.109	.193	.109	.199	.110	-.109	.108	.072	.128	-.147	-.157	1

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Table A3

Subjective and Objective Correlations for the Response Implementation Task Type

Variables	*NASATLX Mental Demand*	*NASATLX Physical Demand*	*NASATLX Temporal Demand*	*NASATLX Effort*	*NASATLX Frustration*	*NASATLX Performance*	*MRQ Auditory Emotional*	*MRQ Auditory Linguistic*	*MRQ Manual*	*MRQ Short Term Memory*	*MRQ Spatial Attentive*
NASATLX Mental Demand	1										
NASATLX Physical Demand	.407**	1									
NASATLX Temporal Demand	.781**	.507**	1								
NASATLX Effort	.760**	.433**	.750**	1							
NASATLX Frustration	.440**	.138	.413**	.551**	1						
NASATLX Performance	.325**	.198	.373**	.337**	.377**	1					
MRQ Auditory Emotional	.306**	.279*	.349**	.309**	.376**	.181	1				
MRQ Auditory Linguistic	.130	.189	.206	.036	-.246*	-.018	-.010	1			
MRQ Manual	.333**	.495**	.250*	.257*	.151	.017	.308**	.152	1		
MRQ Short Term Memory	.179	.097	.246*	.151	-.074	.005	-.053	.583**	.256*	1	
MRQ Spatial Attentive	.240*	.223*	.259*	.112	-.045	-.029	.109	.547**	.440**	.693**	1
MRQ Spatial Concentrative	.128	.156	.173	.127	.110	-.155	.408**	.180	.449**	.306**	.488**
MRQ Spatial Categorical	.097	.217	.231*	.149	-.046	-.082	.295**	.405**	.324**	.438**	.534**
MRQ Spatial Emergent	.087	.173	.173	.096	.036	.013	.225*	.403**	.366**	.471**	.502**
MRQ Spatial Positional	.249*	.127	.262*	.211	-.026	-.061	.114	.418**	.375**	.495**	.589**
MRQ Spatial Quantitative	-.094	.283*	.003	-.016	-.067	.040	.250*	.102	.287**	.137	.317**
MRQ Visual Lexical	.127	.124	.198	.086	-.082	-.047	.044	.444**	.290**	.496**	.555**
MRQ Visual Phonetic	.119	.313**	.213	.101	.008	.189	.200	.502**	.283*	.411**	.475**
MRQ Visual Temporal	-.020	.213	.106	.055	.228*	.064	.519**	.089	.236*	.073	.231*
MRQ Vocal Process	.132	.220*	.122	.083	-.137	-.002	.109	.601**	.390**	.544**	.545**
ISA	.148	.080	.197	.306**	.279*	.224*	.132	.150	.071	.112	.090

Variables	*NASATLX Mental Demand*	*NASATLX Physical Demand*	*NASATLX Temporal Demand*	*NASATLX Effort*	*NASATLX Frustration*	*NASATLX Performance*	*MRQ Auditory Emotional*	*MRQ Auditory Linguistic*	*MRQ Manual*	*MRQ Short Term Memory*	*MRQ Spatial Attentive*
EEG Alpha Left Mean	.145	.062	.079	.165	.085	-.056	.173	.308**	.085	.115	.133
EEG Alpha Right Mean	.107	-.084	.030	.135	.057	-.134	-.102	-.071	-.151	-.129	-.127
EEG Beta Left Mean	.142	.063	.077	.162	.084	-.056	.171	.309**	.085	.116	.133
EEG Beta Right Mean	.107	-.083	.030	.137	.056	-.134	-.100	-.069	-.149	-.126	-.127
EEG Theta Left Mean	.150	.058	.084	.172	.085	-.058	.178	.309**	.089	.115	.135
EEG Theta Right Mean	.108	-.085	.031	.136	.058	-.136	-.102	-.073	-.151	-.130	-.129
TCD Left Mean	-.030	-.130	-.107	-.110	-.055	.024	.011	.181	.270*	.229*	.176
TCD Right Mean	-.120	-.140	-.183	-.052	-.077	-.106	.046	.103	-.027	.160	-.011
fNIR Left Mean	.101	-.076	.004	.120	.247*	.139	.127	-.091	.158	-.051	-.074
fNIR Right Mean	.175	.100	.065	.147	.258*	-.006	.184	-.069	.210	-.134	-.095
ECG HR	.113	.035	.129	.141	.079	-.013	-.001	.074	.076	.158	.238*
ECG HRV	-.001	.117	.090	.006	.058	-.100	.068	.078	.082	.135	.034
ECG IBI	.145	.095	.140	.169	.097	-.105	.044	.013	.132	.148	.133

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Variables	*MRQ Spatial Concentrative*	*MRQ Spatial Categorical*	*MRQ Spatial Emergent*	*MRQ Spatial Positional*	*MRQ Spatial Quantitative*	*MRQ Visual Lexical*	*MRQ Visual Phonetic*	*MRQ Visual Temporal*	*MRQ Vocal Process*	*ISA*	*EEG Alpha Left Mean*
NASATLX Mental Demand											
NASATLX Physical Demand											
NASATLX Temporal Demand											
NASATLX Effort											
NASATLX Frustration											
NASATLX Performance											
MRQ Auditory Emotional											
MRQ Auditory Linguistic											
MRQ Manual											
MRQ Short Term Memory											
MRQ Spatial Attentive											
MRQ Spatial Concentrative	1										
MRQ Spatial Categorical	.590**	1									
MRQ Spatial Emergent	.456**	.661**	1								
MRQ Spatial Positional	.427**	.625**	.605**	1							
MRQ Spatial Quantitative	.541**	.491**	.406**	.396**	1						
MRQ Visual Lexical	.288**	.428**	.323**	.550**	.298**	1					
MRQ Visual Phonetic	.407**	.412**	.366**	.414**	.510**	.605**	1				
MRQ Visual Temporal	.573**	.393**	.405**	.262*	.445**	.154	.424**	1			
MRQ Vocal Process	.224*	.316**	.340**	.375**	.203	.518**	.429**	.136	1		
ISA	-.018	.074	.038	.104	-.006	.162	.222*	.020	.110	1	

Variables	*MRQ Spatial Concentrative*	*MRQ Spatial Categorical*	*MRQ Spatial Emergent*	*MRQ Spatial Positional*	*MRQ Spatial Quantitative*	*MRQ Visual Lexical*	*MRQ Visual Phonetic*	*MRQ Visual Temporal*	*MRQ Vocal Process*	*ISA*	*EEG Alpha Left Mean*
EEG Alpha Left Mean	.073	.018	.022	-.054	-.125	.056	.165	.219	.251*	.078	1
EEG Alpha Right Mean	-.118	.011	.016	.159	-.106	-.068	-.150	-.075	-.165	-.026	-.040
EEG Beta Left Mean	.072	.018	.022	-.055	-.122	.056	.167	.219	.252*	.076	1
EEG Beta Right Mean	-.115	.016	.015	.160	-.108	-.066	-.147	-.072	-.163	-.023	-.039
EEG Theta Left Mean	.077	.023	.025	-.051	-.134	.057	.157	.219	.253*	.082	.999**
EEG Theta Right Mean	-.117	.013	.017	.160	-.106	-.069	-.150	-.075	-.166	-.028	-.042
TCD Left Mean	.177	.145	.049	.158	.088	.306**	.065	.030	.431**	.044	-.019
TCD Right Mean	.140	-.002	-.003	-.043	-.143	-.176	-.137	.092	.020	-.116	.149
fNIR Left Mean	-.032	-.097	-.019	.032	-.071	.223*	-.005	-.165	-.007	.151	-.069
fNIR Right Mean	-.024	-.256*	-.160	-.101	-.269*	-.129	-.236*	-.182	-.018	-.020	-.065
ECG HR	.020	.091	.101	.076	-.006	.113	.101	-.054	.089	-.023	.091
ECG HRV	-.013	.164	.140	.023	.109	.038	.007	.114	.142	-.175	.278*
ECG IBI	.123	.104	.093	-.033	-.091	.034	-.086	.070	.052	-.158	.232*

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Variables	*EEG Alpha Right Mean*	*EEG Beta Left Mean*	*EEG Beta Right Mean*	*EEG Theta Left Mean*	*EEG Theta Right Mean*	*TCD Left Mean*	*TCD Right Mean*	*fNIR Left Mean*	*fNIR Right Mean*	*ECG HR*	*ECG HRV*	*ECG IBI*
NASATLX Mental Demand												
NASATLX Physical Demand												
NASATLX Temporal Demand												
NASATLX Effort												
NASATLX Frustration												
NASATLX Performance												
MRQ Auditory Emotional												
MRQ Auditory Linguistic												
MRQ Manual												
MRQ Short Term Memory												
MRQ Spatial Attentive												
MRQ Spatial Concentrative												
MRQ Spatial Categorical												
MRQ Spatial Emergent												
MRQ Spatial Positional												
MRQ Spatial Quantitative												
MRQ Visual Lexical												
MRQ Visual Phonetic												
MRQ Visual Temporal												
MRQ Vocal Process												
ISA												

Variables	*EEG Alpha Right Mean*	*EEG Beta Left Mean*	*EEG Beta Right Mean*	*EEG Theta Left Mean*	*EEG Theta Right Mean*	*TCD Left Mean*	*TCD Right Mean*	*fNIR Left Mean*	*fNIR Right Mean*	*ECG HR*	*ECG HRV*	*ECG IBI*
EEG Alpha Left Mean												
EEG Alpha Right Mean	1											
EEG Beta Left Mean	-.039	1										
EEG Beta Right Mean	1	-.038	1									
EEG Theta Left Mean	-.041	.999**	-.039	1								
EEG Theta Right Mean	1	-.041	.999**	-.042	1							
TCD Left Mean	-.220	-.020	-.218	-.017	-.219	1						
TCD Right Mean	.030	.147	.030	.151	.031	.191	1					
fNIR Left Mean	-.198	-.070	-.199	-.066	-.197	.214	-.077	1				
fNIR Right Mean	-.140	-.068	-.141	-.059	-.139	.189	.136	.664**	1			
ECG HR	-.067	.091	-.069	.092	-.067	-.037	-.053	-.130	-.214	1		
ECG HRV	.100	.280*	.100	.278*	.096	-.066	-.150	-.075	-.202	.392**	1	
ECG IBI	-.036	.231*	-.034	.235*	-.037	-.045	.039	-.105	-.153	.516**	.450**	1

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

APPENDIX B:
INFORMED CONSENT

University of Central Florida

Investigating Measures of Workload for Nuclear Power Plant Operators

Informed Consent

Principal Investigator(s):	*Lauren Reinerman-Jones, Ph.D.*
Co-Investigator(s):	*Joseph Mercado, M.S.* *Rebecca Leis, M.S.* *Peter A. Hancock, Ph.D.*
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.

University of Central Florida IRB
IRB Number: SBE-13-09210
IRB EXPIRATION DATE: 03/20/2014

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

What you will be asked to do in the study:

You will be doing task 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. 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 5 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 5 hours performing experimental tasks, for which they may elect to receive course credit or cash payment of $8/hr. for the amount of time they participate. Maximum course credit will be 5 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 C:
RESTRICTIONS CHECKLIST

Group:	
Participant number:	
Date:	
Start time:	

Check the box if the answer is yes

Are you less than 18 years old?	☐
Are you greater than 40 years old?	☐
Have you had any caffeine in the last 2 hours?	☐
Have you had any nicotine in the last 2 hours?	☐
Have you had any Alcohol in the last 24 hours?	☐
Have you had any aspirin, tylenol, or similar medications in the last 24 hours?	☐
Have you had any antihistamines or decongestants in the last 24 hours?	☐
Have you had any sedatives or tranquilizers in the last 24 hours?	☐
Have you had any anti-psychotics or anti-depressants in the last 24 hours?	☐
Based on your current knowledge, are you pregnant?	☐
Do you have any metal plates in your head?	☐
Do you lack normal or corrected to normal vision?	☐
Are you colorblind?	☐
Do you have a history of epilepsy or seizures?	☐
Is your hair wet?	☐
Do you have any impairment of your dominant arm or hand?	☐
Are you right handed?	☐

Check the box if you use your right hand

Which hand do you use to write with?	☐
Which hand do you use to throw a ball?	☐
Which hand do you hold a toothbrush with?	☐
Which hand holds a knife when you cut things?	☐
Which hand holds a hammer when you nail things?	☐

APPENDIX D:
DEMOGRAPHICS QUESTIONNAIRE

Demographics Questionnaire

Participant # _______ Age ______ Major _______________ Date ___________ Gender ___

1. What is the highest level of education you have had?
Less than 4 yrs of college ____ Completed 4 yrs of college ____ Other ____

2. When did you use computers in your education? (*Circle all that apply*)

Grade School	Jr. High	High School
Technical School	College	Did Not Use

3. Where do you currently use a computer? (*Circle all that apply*)
Home Work Library Other________ Do Not Use

4. How many hours per day do you use a computer? ___________

5. For each of the following questions, circle the response that best describes you.

How often do you:	
Use a mouse?	Daily, Weekly, Monthly, Once every few months, Rarely, Never
Use a joystick?	Daily, Weekly, Monthly, Once every few months, Rarely, Never
Use a touch screen?	Daily, Weekly, Monthly, Once every few months, Rarely, Never
Use icon-based programs/software?	Daily, Weekly, Monthly, Once every few months, Rarely, Never
Use programs/software with pull-down menus?	Daily, Weekly, Monthly, Once every few months, Rarely, Never
Use graphics/drawing features in software packages?	Daily, Weekly, Monthly, Once every few months, Rarely, Never
Use E-mail?	Daily, Weekly, Monthly, Once every few months, Rarely, Never
Operate a radio controlled vehicle (car, boat, or plane)?	Daily, Weekly, Monthly, Once every few months, Rarely, Never
Play computer/video games?	Daily, Weekly, Monthly, Once every few months, Rarely, Never

6. Which type(s) of computer/video games do you most often play if you play at least once every few months?

7. Which of the following best describes your expertise with computers? (check √ one)
_____ Novice
_____ Good with one type of software package (such as word processing or slides)
_____ Good with several software packages
_____ Can program in one language and use several software packages
_____ Can program in several languages and use several software packages

8. How many hours per day do you watch television? ________

9. How many hours per day do you spend reading? __________

10. Are you in your usual state of health physically? YES NO
If NO, please briefly explain:

11. How many hours of sleep did you get last night? ______ hours

12. What is your occupation? _____________

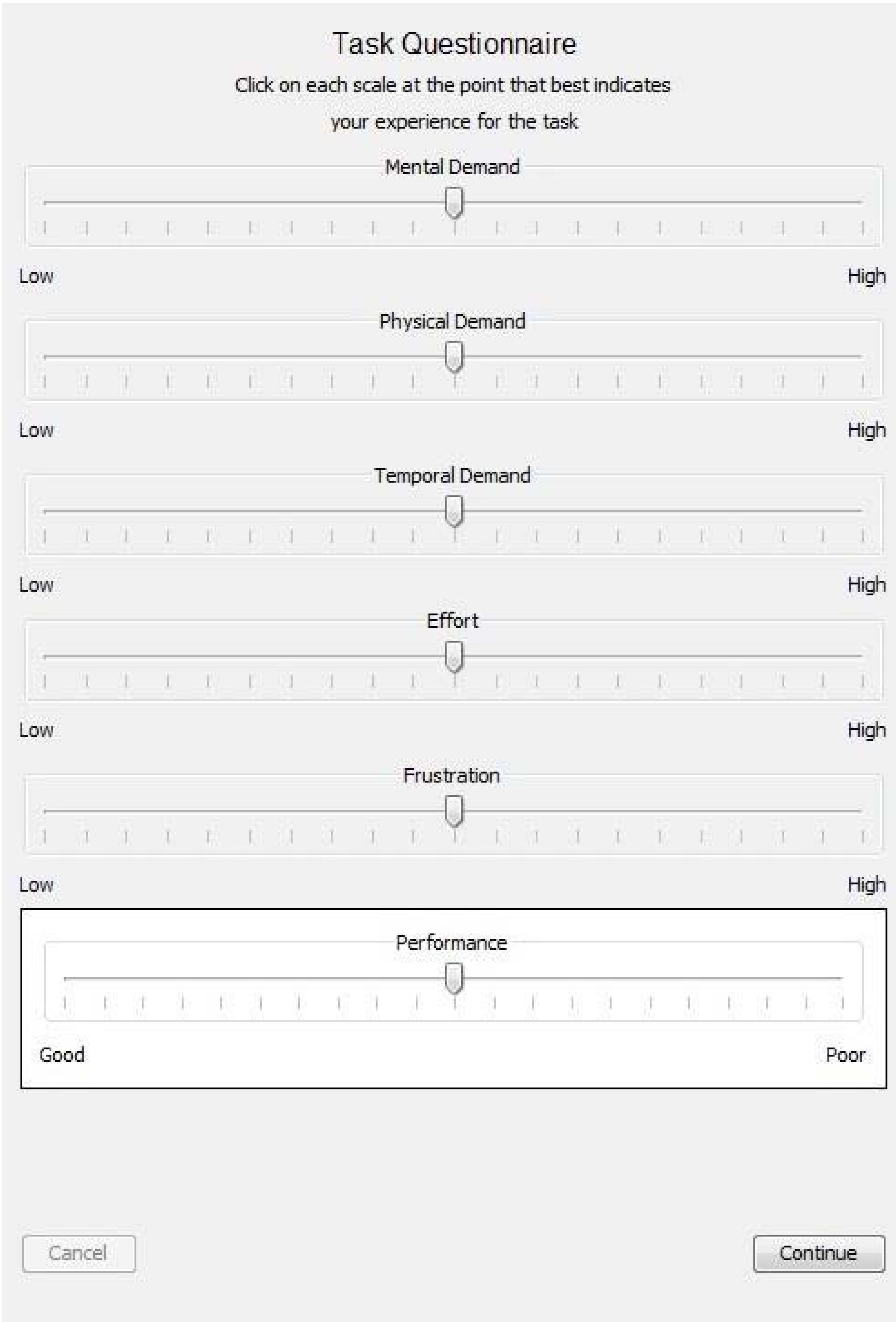
13. How often do you feel eye strain?

0	1	2	3	4	5
Not at all	Mildly		Average		Highly

14. During an average work day, do you feel that you focus on near objects (about 2 meters away) more than objects that are far away (6 meters or more)?

1	2	3	4	5
Strongly disagree		Agree		Strongly agree

APPENDIX E:
NASA-TASK LOAD INDEX



Task Questionnaire

Click on each scale at the point that best indicates
your experience for the task

Mental Demand

Low High

Physical Demand

Low High

Temporal Demand

Low High

Effort

Low High

Frustration

Low High

Performance

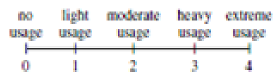
Good Poor

Cancel Continue

APPENDIX F:
MULTIPLE RESOURCE QUESTIONNAIRE

MULTIPLE RESOURCES QUESTIONNAIRE for task ________________________

The purpose of this questionnaire is to characterize the nature of the mental processes used in the task with which you have become familiar. Below are the names and descriptions of several mental processes. Please read each carefully so that you understand the nature of the process. Then rate the task on the extent to which it uses each process, using the following scale.



Important:

All parts of a process definition should be satisfied for it to be judged as having been used. For example, recognizing geometric figures presented visually should **not** lead you to judge that the "Tactile figural" process was used, just because figures were involved. For that process to be used, figures would need to be processed tactilely (i.e., using the sense of touch).

Please judge the task as a **whole**, averaged over the time you performed it. If a certain process was used at one point in the task and not at another, your rating should **not** reflect "peak usage" but should instead reflect **average** usage over the entire length of the task.

Auditory emotional process – Required judgments of emotion (e.g., tone of voice or musical mood) presented through the sense of hearing.

Auditory linguistic process – Required recognition of words, syllables, or other verbal parts of speech presented through the sense of hearing.

Facial figural process – Required recognition of faces, or of the emotions shown on faces, presented through the sense of vision.

Facial motive process – Required movement of your own face muscles, unconnected to speech or the expression of emotion.

Manual process – Required movement of the arms, hands, and/or fingers.

Short-term memory process – Required remembering of information for a period of time ranging from a couple of seconds to half a minute.

Spatial attentive process – Required focusing of attention on a location, using the sense of vision.

Spatial categorical process – Required judgment of simple left-versus-right or up-versus down relationships, without consideration of precise location, using the sense of vision.

Spatial concentrative process – Required judgment of how tightly spaced are numerous visual objects or forms.

Spatial emergent process – Required "picking out" of a form or object from a highly cluttered or confusing background, using the sense of vision.

Spatial positional process - Required recognition of a precise location as differing from other locations, using the sense of vision.

Spatial quantitative process – Required judgment of numerical quantity based on a nonverbal, nondigital representation (for example, bar graphs or small clusters of items), using the sense of vision.

Tactile figural process – Required recognition or judgment of shapes (figures), using the sense of touch.

Visual lexical process – Required recognition of words, letters, or digits, using the sense of vision.

Visual phonetic process – Required detailed analysis of the sound of words, letters, or digits, presented using the sense of vision.

Visual temporal process – Required judgment of time intervals, or of the timing of events, using the sense of vision.

Vocal process – Required use of your voice.

APPENDIX G:
IRB APPROVAL LETTER

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

Page 1 of 2

REFERENCES

Abich, J. IV. (2013). *Investigating the universality and comprehensive ability of measures to assess the state of workload*. (Doctoral Dissertation). University of Central Florida.

Ayaz, H., Shewokis, P. A, Bunce, S., Izzetoglu, K., Willems, B., & Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *NeuroImage, 59*(1), 36–47.

Ayaz, H., Willems, B., Bunce, S., Shewokis, P. A., Hah, S., Deshmukh, A. R., & Onaral, B. (2010). Cognitive workload assessment of air traffic controllers using optical brain imaging sensors. In T. Marek, W. Karwowski, & V. Rice (Eds.), *Advances in Understanding Human Performance: Neuroergonomics, Human Factors Design, and Special Populations* (pp. 21–32). Taylor & Francis Group.

Baker, T. L., Campbell, S. C., Linder, K. D., & Moore-Ede, M. C. (1990). *Control-room operator alertness and performance in nuclear power plants* (No. EPRI-NP-6748). Environmental Protection Agency, Washington, DC U.S. Circadian Technologies, Inc., Boston, MA

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

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

Boles, D. B. (1996). Factor analysis and the cerebral hemispheres: “unlocalized” functions. *Neuropsychologia*, *34*(7), 723–36.

Boles, D. B. (2002). Lateralized spatial processes and their lexical implications. *Neuropsychologia*, *40*(12), 2125–35.

Boles, D. B., & Adair, L. P. (2001). The Multiple Resources Questionnaire (MRQ). *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *45*(25), 1790–1794.

Brannick, M. T., Salas, E., & Prince, C. W. (Eds.). (1997). *Team performance assessment and measurement: Theory, methods, and applications.* Psychology Press.

Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology*, *42*(3), 361–377.

Brouwer, A. M., Hogervorst, M. A., Van Erp, J. B., Heffelaar, T., Zimmerman, P. H., &

Oostenveld, R. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of Neural Engineering*, *9*(4), 045008.

Brown, T. J., Conrad, S. H., & Beyeler, W. E. (2012). Complex Adaptive Systems Engineering-improving our understanding of complex systems and reducing their risk.

Cain, B. (2007). *A review of the mental workload literature*. Defense Research and Development, Toronto, Canada

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

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurements, 33,* 107-112

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Eggemeier, F. T., Wilson, G. F., Kramer, A. F., & Damos, D. (1991). Workload assessment in multi-task environments. In D. L. Damos (Ed.), *Multiple-task performance* (pp. 207-216). London, England: Taylor & Francis.

Erwin, M. C. (2013, April). *Intelligence, Surveillance, and Reconnaissance (ISR) Acquisition: Issues for Congress.* Library of Congress Washington DC, U.S. Congressional Research Service.

Friedenberg, J., & Silverman, G. (2006). *Cognitive science: An introduction to the study of mind*. Sage.

Gratton, G., Fabiani, M., 2008. Optical imaging of brain function. In: Parasuraman, R., Rizzo, M. (Eds.), *Neuroergonomics: The Brain at Work.* Oxford University Press, Cambridge, MA, pp. 65–81

Gopher, D., & Donchin, E. (1986). Workload – An examination of the concept. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 41-1-41-49). New York, NY: John Wiley & Sons, Inc.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.

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

Guznov, S., Reinerman-Jones, L., & Marble, J. (2012). Applicability of situation awareness and workload metrics for use in assessing nuclear power plant designs. In K. M. Stanney & K. S. Hale (Eds.), Advances in Cognitive Engineering and Neuroergonomics (pp. 91-98). Boca Raton, FL: Taylor & Francis Group.

Hancock, P. A. (1989). A dynamic model of stress and sustained attention. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 31*(5), 519-537.

Hancock, P. A., Mercado, J. E., Merlo, J., & Van Erp, J. B. (2013). Improving target detection in visual search through the augmenting multi-sensory cues. *Ergonomics, 56*(5), 729-738.

Hancock, P. A., & Verwey, W. B. (1997). Fatigue, workload and adaptive driver systems. *Accident Analysis & Prevention, 29*(4), 495-506.

Hancock, P. A., and Warm, J. S. (1989). A dynamic model of stress in sustained attention. *Human Factors, 31*, 519-537.

Hart, C. S. (2010). Assessing the Impact of Low Workload in Supervisory Control of Networked Unmanned Vehicles. Cambridge, MA: Massachusetts Institute of Technology.

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*(9), 904-908.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), Human Mental Workload (pp. 139-184). North-Holland: Elsevier Science Publishers.

Hart, S. G., & Wickens, C. D. (1990). Workload assessment and prediction. In H. R. Booher (Ed.), MANPRINT: An emerging technology. Advanced concepts for integrating people, machines, and organizations (pp. 257–300). New York, NY: Van Nostrand Reinhold.

Heeger, D. (1997). Signal Detection Theory. Department of Psychology, Standford University.

Henelius, A., Hirvonen, K., Holm, A., Korpela, J., & Muller, K. (2009). Mental workload classification using heart rate metrics. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2–6.

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science, 143*, 1190–1192.

Hinds, P. J. (1999). The curse of expertise: The effects of expertise and debiasing methods on prediction of novice performance. *Journal of Experimental Psychology: Applied, 5*(2), 205.

Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology,* *45*(1-3), 73-93.

Hockey, G. R., & Tattersall, A. J. (1995). Level of operator control and changes in heart rate variability during simulated flight maintenance. *Human Factors, 37*(4), 682–698.

Huey, B. M., & Wickens, C. D. (Eds.). (1993). Workload transition: Implications for individual and team performance. National Academies Press.

Hulbert, T. (1989). A comparison of the 'NASA-TLX' and 'ISA' subjective workload rating techniques. Internal Report, Civil Aviation Authority Air Traffic Control Evaluation Unit, Bournemouth, UK.

Hwang, S. L., Yau, Y. J., Lin, Y. T., Chen, J. H., Huang, T. H., Yenn, T. C., & Hsu, C. C. (2008). Predicting work performance in nuclear power plants. *Safety Science, 46*(7), 1115-1124.

Johannsen, G. (1979). Workload and workload measurement. In N. Moray (Ed.), Mental workload (pp. 3–12). New York, NY: Plenum Press.

Jordan, C.S. (1992). Experimental study of the effect of an instantaneous self-assessment workload recorder on task performance. Defense Research agency Technical Memorandum DRA TM (CADS) 92011, DRA, Portsdown, Hants.

Jorna, P. G. A. M. (1993). Heart-rate and workload variations in actual and simulated flight. *Ergonomics, 36*(9), 1043–1054.

Kahneman, D. (1973). Attention and effort. Englewood Cliffs, NJ: Prentice Hall.

Kahneman, D., & Teisman, A. (1984). Changing views of attention and automaticity. In

Kahneman, D., Teisman, A., & Gibbs, B. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24,* 175–219.

Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. In D. L. Damos (Ed.), Multiple-task performance. London, England: Taylor & Francis Group.

Kramer, A. F., Sirevaag, E. J., & Braune, R. (1987). A psychophysiological assessment of operator workload during simulated flight missions. *Human Factors, 29*(2), 145-160.

Lin, C. J., Hsieh, T. L., Tsai, P. J., Yang, C. W., & Yenn, T. C. (2011). Development of a team workload assessment technique for the main control room of advanced nuclear power plants. *Human Factors and Ergonomics in Manufacturing & Service Industries, 21*(4), 397-411.

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

Matthews, G., & Campbell, S. E. (1999, September). Individual differences in stress response and working memory. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 43*(11), 634-638.

Matthews, G., Campbell, S., Falconer, S., Joyner, L., Huggins, J., Gilliland, K., Grier, R., et al. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. *Emotion*, *2*(4), 315-340.

Matthews, G., Davies, D. R., Westerman, S. J., & Stammers, R. B. (2000). Human performance: Cognition, stress, and individual differences. Philadelphia, PA: Taylor & Francis Group.

Mercado, J. E., White, T. L., Sanders, T., Wright, J., & Hancock, P. A. (2013). Performance Effects of Imperfect Cross-Modal Sensory Cueing in a Target Detection Simulation. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, 1548512913509908.

Meshkati, N., & Hancock, P. A. (Eds.). (2011). *Human mental workload*. New York, NY: Elsevier.

Meshkati, N., & Loewenthal, A. (1988). An eclectic and critical review of four primary mental workload assessment methods: A guide for developing a comprehensive model. *Advances in Psychology*, *52*, 251-267.

Metzger, U., & Parasuraman, R. (2005). Automation in future air traffic management: Effects of decision aid reliability on controller performance and mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *47*(1), 35-49.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, *63*, 81-97.

Miller, S. (2001). Literature review: Workload Measures. Iowa City, IA: National Advanced Driving Simulator Technical Report N01-006.

Moray, N. (1967). Where is capacity limited? A survey and a model. *Acta Psychologica*, *27*, 84–92.

Moray, N. (1979). Models and measures of mental workload. Mental workload: Its theory and measurement. New York, NY: Plenum Press.

Muckler, F. A. (1992). Selecting performance measures: Objective versus subjective measurement. *Human Factors*, *34*(4), 441–455.

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review*, *86*(3), 214.

O'Donnell, R.D. Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff & J. P. Thomas (Eds.), Handbook of perception and human performance Vol. 2: Cognitive processes and performance (2nd ed., pp. 1–49). Oxford, England: John Wiley & Sons.

O'Hara, J. M., & Higgins, J. C. (2010). Human-System interfaces to automatic systems: Review guidance and technical bases. Human Factors of advanced reactors (NRC JCN Y-6529) BNL Tech Report No BNL91017-2010.

OHara, J., Higgins, J., Brown, W., & Fink, R. (2008). *Human Factors Considerations in New Nuclear Power Plants: Detailed Analysis* (No. BNL--79947-2008). Brookhaven National Laboratory.

Paas, F. G., & Van Merriënboer, J. J. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *35*(4), 737-743.

Paas, F. G;, & Van Merriënboer, J. J. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, *6*(4), 351–371.

Pearson, K. (1911). On a correction of needful in the case of correlation ratio. *Biometrika 8,* 254-256

Putze, F., Jarvis, J. P., & Schultz, T. (2010, August). Multimodal recognition of cognitive workload for multitasking in the car. In *Pattern Recognition (ICPR), 2010 20th International Conference on* (pp. 3748-3751). IEEE.

Rasmussen, J. (1979). Reflection on the concept of workload. In N. Moray (Ed.), Mental workload (pp. 29–40). New York, NY: Plenum Press.

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

Reinerman-Jones, L.E., Cosenzo, K., & Nicholson, D. (2010). Subjective and objective measures of operator state in automated systems. In T. Marek, W. Karwowski, & V. Rice (Eds.), Advances in understanding human performance: Neuroergonomics, human factors design, and special populations.

Reinerman-Jones, L., Guznov, S., Mercado, J., & D'Agostino, A. (2013). Developing Methodology for Experimentation Using a Nuclear Power Plant Simulator. In *Foundations of Augmented Cognition* (pp. 181-188). Springer Berlin Heidelberg.

Reinerman-Jones, L.E., Guznov, S., Tyson, J., & D'Agostino, A. (2012). Workload,Situation Awareness, and Teamwork. (No. NUREG-under review) Nuclear Regulatory Commission, Washington D.C.

Reinerman-Jones, L. E., Matthews, G., Langheim, L. K., & Warm, J. S. (2011). Selection for vigilance assignments: A review and proposed new direction. *Theoretical Issues in Ergonomics Science, 12*(4), 273-296.

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. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 50*(12), 1254-1258.

Rensink, R. A. (1997). How much of a scene is seen? The role of attention in scene perception. *Investigative Ophthalmology and Visual Science*, 38.

Rose, S. P. (1981). From causations to translations: What biochemists can contribute to the study of behavior. In *Perspectives in ethology* (pp. 157-177). Springer US.

See, J. E., Howe, S. R., Warm, J. S., & Dember, W. N. (1995). Meta-analysis of the sensitivity decrement in vigilance. *Psychological Bulletin, 117*(2), 230-249.

Simons, D. J. (1996). Out of sight, out of mind: When object representations fail. *Psychological Science, 7*, 301–305.

Stanton, N. A., Salmon, P. M., Walker, G. H., Baber, C., & Jenkins, D. P. (2005). *A practical guide for engineering and design*. Burlington, VT: Ashgate.

Szalma, J. L., Warm, J. S., Matthews, G., Dember, W. N., Weiler, E. M., Meier, A., & Eggemeier, F. T. (2004). Effects of sensory modality and task duration on performance, workload, and stress in sustained attention. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(2), 219-233.

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

Taylor, G. S. (2012). *Comparing types of adaptive automation within a multi-tasking environment.* (Doctoral Dissertation). University of Central Florida.

Taylor, G., Reinerman-Jones, L., Cosenzo, K., & Nicholson, D. (2010). Comparison of multiple physiological sensors to classify operator state in adaptive automation systems. *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *54*(3), 195-199

Tripp, L. D., & Warm, J. S. (2007). Transcranial Doppler sonography. In R. Parasuraman & M.

Rizzo (Eds.*), Neuroergonomics: The brain at work* (pp. 82–94). New York, NY: Oxford University Press.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*(2), 127-154.

United States Regulatory Commission (2013, December 13). Power Reactors. *NRC.gov*. Retrieved February 05, 2014, from http://www.nrc.gov/reactors/power.html

Veltman, J. A., & Gaillard, A. W. K. (1996). Physiological indices of workload in a simulated flight task. *Biological Psychology, 42*, 323–342.

Warm, J. S., Dember, W. N., & Hancock, P.A. (1996). Vigilance and workload in automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 183–200). Hillsdale, NJ: Erlbaum.

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 433-441.

Wickens, C. D. (1976). The effects of divided attention on information processing and manual tracking. *Journal of Experimental Psychology: Human Perception and Performance Processing in Manual Tracking*, *2*(1), 1–13.

Wickens, C. D. (1984). *Engineering psychology and human performance*. New York, NY: Harper Collins.

Wickens, C. D. (1992). *Engineering psychology and human performance* (2nd ed.). New York, NY: Harper Collins.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors, 50*(3), 449–455.

Wickens, C. D., & Hollands, J. G. (2000*). Engineering psychology and human performance.* Upper Saddle Reiver, New Jersey: Prentice Hall.

Wierwille, W. W. (1988). Important remaining issues in mental workload estimation. *Advances in Psychology*, *52*, 315-327.

Wilson, G. F. (2002). An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures. *The International Journal of Aviation Psychology*, *12*(1), 3–18.

Wilson, G. F., Fullenkamp, P., & Davis, I. (1994). Evoked-potential, cardiac, blink, and respiration measures of pilot workload in air-to-ground missions. *Aviation Space and*

Environmental Medicine, 65(2), 100–105.

Wilson, G. F., & O'Donnell, R. D. (1988). Measurement of operator workload with the neuropsychological workload test battery. *Advances in Psychology, 52*, 63-100.

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