ABSTRACT

ON THE MEASUREMENT OF SITUATION AWARENESS

IN PETROCHEMICAL REFINING

By

Hector I. Silva

August 2015

The petrochemical field is an industry seeking to increase efficiency, improve safety of workers, and lessen environmental impacts. One way to improve the performance of operators is to investigate their situation awareness (SA). Research has shown that SA is a predictor of performance. However, there is little consensus on how to measure SA. This study investigated two prominent techniques for measuring SA: the Situation Present Assessment Method (SPAM) and the Situation Awareness Global Assessment Technique (SAGAT). These two techniques were examined for their psychometric properties in assessing SA among operators. The results of this investigation showed that probe-type SA techniques can be used to assess SA in this field. This especially applies to the SPAM technique, which was shown to predict performance, not intrude, and was preferred by a majority of operators.

ON THE MEASUREMENT OF SITUATION AWARENESS

IN PETROCHEMICAL REFINING

A THESIS

Presented to the Department of Psychology

California State University, Long Beach

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Psychology

Option in Human Factors

Committee Members:

Dan Chiappe, Ph.D. (Chair)
Kim-Phuong L. Vu, Ph.D.
Mary A. Burress, Ph.D.

College Designee:

Dean Amy Bippus, Ph.D.

By Hector I. Silva

B.A., 2013, California State University, Long Beach

August 2015

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

TABLE Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In the early afternoon of March 23, 2005, operation crews were restarting the isomerization process unit of a large refinery in the United States when a fire ignited and caused a major explosion that killed 15 people, injured another 180, and damaged communities three-quarters of a mile away (U.S. Chemical Safety and Hazard Investigation Board, 2007). The isomerization chemical process increases the octane in certain hydrocarbons that are later used in the production of gasoline. The unit had been disabled in order for maintenance to be completed on the unit. During start up, a main control valve designed to send the liquid hydrocarbons from the main tower of the isomerization unit to storage tanks had been closed, leaving the isomerization tower to overfill for more than 3 hours. This caused a flooding of the tower and automatic relief valves to open and release liquid. As the flammable liquid overfilled, it began to evaporate and form a gas cloud. As the gas cloud grew larger, investigators suspect that the cloud reached an idling truck which served as the ignition source.

An investigation by Saleh, Haga, Favaro, and Bakolas (2014) cited a lack of situation awareness (SA) on the part of control board operators as a major contributing factor in this incident. Incidents such as these have massive consequences for the environment, the safety of workers and surrounding communities, and lead to large monetary losses that, in this case, exceeded $1 billion. Perceptions of the companies involved in these incidents and the level of concern regarding personal safety have also been shown to change negatively after such incidents (Cutchin, Martin, Owen, & Goodwin, 2008). Analogous industries such as the nuclear industry have also been scrutinized in the past for disastrous incidents and have begun to be regulated by government agencies including the U.S. Nuclear Regulatory Commission as a result (2012). Parts of these regulations include Human Factors guidelines and principles to be followed throughout the system. It is likely that similar regulations will soon be implemented in petrochemical refining and other high-stakes industries (Department of Industrial Relations Division of Occupational Safety and Health, 2014).

Having a proactive and diagnostic system of human factors evaluation could help the petrochemical field to move forward towards improving safety, efficiency, and environmental impacts. One human factors psychology construct that could be assessed in the petrochemical industry is SA. The construct of SA has been a topic of discussion among Human Factors/Ergonomists researchers for decades (Chiappe, Strybel, & Vu, 2012; Durso et al., 1999; Endsley, 1995b; Parasuraman, Sheridan, & Wickens, 2008). SA is a psychological construct that, in simple terms, involves “knowing what’s going on” as an operator is managing a complex, dynamic system (Chiappe et al., 2012; Endsley, 1995b). One key reason SA is an important topic of discussion is because of its contribution to human performance in complex tasks (Durso et al, 1999; Durso, Bleckley, & Dattel, 2006; Endsley, 1995b). Indeed, it can serve as an important bellwether, as system operators may suffer from decreases in SA before there are any apparent decrements in performance (Durso & Dattel, 2004). Because of this, several industries have attempted to measure SA in the human agents that support those systems (Endsley, 2012).

In the case of petrochemical refining, issues regarding SA are particularly important because console operators are often placed in monitoring roles as a part of highly automated systems. Past research has indicated that these passive operating roles can lead to the "out of the loop" performance problems (Endsley & Kaber, 1999; Kaber & Endsley, 1997, 2004). As such, SA can be a major factor affecting a console operator's ability to manage off-nominal situations efficiently and, most importantly, safely. The petrochemical field has seen an increase in automation as a means to lower human error. However, automation has to be implemented carefully because it can lead to unforeseen, unwanted outcomes (Kaber & Endsley, 1997). Console operators may suffer negative consequences of automation such as: a failure to detect changes in critical process parameters, over reliance on automation, a decay of skill or competency, and decreases in SA (Kaber & Endsley, 1997). In the case of petrochemical refining, console operators are typically in a state of under-load, where not much human intervention is required of them as they are simply monitoring automation. This under-loaded state has been cited as detrimental in both nominal and off-nominal situations (Kaber & Endsley, 2004). In the latter, console operators are typically thrust back into the task to diagnose what has gone wrong, often facing steep increases in workload. The efficacy with which console operators diagnose and manage an off-nominal event is highly dependent on their level of SA when the event occurs. Therefore, it is important to be able to measure the SA of console operators to understand how their SA fluctuates due to such factors as changes in workload.

In sum, the petrochemical industry would benefit greatly by taking into account SA as a means of assisting in ultimately reducing negative consequences of automation, minimizing incidents from occurring, and optimizing its processes. The following is a review of SA and the qualifications for an effective measurement technique of the construct that can be potentially used in petrochemical refining.

Theoretical Underpinning of Situation Awareness

Despite the widespread agreement that SA is an important factor influencing performance, there is still disagreement amongst researchers as to the theoretical underpinnings of the construct and how best to measure it (Chiappe et al., 2012; Endsley, 1995b; Stanton et al., 2006). One area of contention has to do with whether SA refers to the process of acquiring and maintaining awareness of the situation or the product that is a consequence of those processes. Another point of discussion is whether SA is strictly an internalized representational state, or if it is a property that relies fundamentally on an interaction between the agent and the environment, with external representations also playing a crucial role (Chiappe et al., 2012; Endsley, 1995b; Stanton et al., 2006).

Given the role that SA plays in predicting performance, it is important to design system displays in such a way that they facilitate the acquisition and maintenance of SA. Doing so, however, requires tools to measure SA to ensure that the displays are actually having the intended effect and are superior to alternatives. As we will see below, however, selecting a suitable metric is complicated by the fact that many techniques presuppose conflicting theories of SA (Lau, Jamieson, & Skraaning, 2013; Salmon et al., 2009). Although it has been noted that most of these theoretical models have at least some merits, it is unlikely that a definition of SA will be universally accepted in the near future (Salmon et al., 2009).

<u>Three-Level Model of Situation Awareness</u>

As mentioned above, theories of SA differ in whether SA is viewed as something that is strictly held in an individual mind or whether it also relies on external factors and whether it is a product or process. The most prominent argument for SA as a product existing inside the individual mind of an operator comes from Endsley (1995b). Endsley (1995b) argued that SA is composed of three levels. Level 1 SA is the perception of elements in the environment. It represents elements of information that are not yet interpreted or integrated. Level 2 SA is the comprehension of the perceivable elements. It is the integration of individual pieces of information to form more holistic patterns of awareness where elements are processed in light of current goals. Level 3 SA is the projection of the future status of the environment based on the understanding (Level 2 SA) of the individual environmental elements (Level 1 SA). An example with console operators would be if they were to perceive an off-nominal alarm of a certain process parameter, understand what the alarm means or how it affects their goals to know if any action is required, and predicting how their action or inaction would affect the process parameter in the near future.

These levels produce what Endsley (1995b) called a "situational model," which is a detailed, stable representation of the situation maintained in internal memory. According to this model, the processes used to acquire SA are separate from the construct itself, the former being referred to as "situational assessment." Situational assessment depends on cognitive processes such as perception, working memory, short term memory, and the mental models and schemata in long term memory. Although Endsley (2012) stated that SA occurs in working memory, she also claims that long term memory plays a part. This is because mental models and schemata stored in long term memory can be used to rapidly and efficiently encode perceptual information. These long term memory structures also enhance comprehension and projection, thus facilitating the process of acquiring SA.

Endsley (1995a) has proposed a technique to measure SA that presupposes the three-level model of SA. The Situation Awareness Global Assessment Technique (SAGAT) is an offline probe technique where a battery of probe questions regarding the situation are presented to participants during a simulation freeze that occurs at random times throughout a scenario. The simulation is paused and all information-bearing displays are blanked. This is consistent with Endsley's (1995b) notion that SA is a detailed internalized representation of the situation. SA is measured as the accuracy of responses to the probe questions administered. The probe questions tap each of the three levels of SA. They are constructed using a task analysis in conjunction with SME consultation.

Viewing SA as a product, however, does not elucidate the processes by which SA is acquired. Durso and Sethumadhavan (2008) stated that understanding the process of acquiring SA is important for developing training programs that facilitate the acquisition of SA and also interventions for preventing the loss of SA, which might be of particular importance for the petrochemical field. This may be especially important for console operators at refinery plants whose job is primarily to monitor processes that rely heavily on automated technology.

Distributed and Situated Situation Awareness

Other alternatives to Endsley's (1995b) Three-Level Model theory have been proposed to explain SA as a process. One such theory is Stanton et al.'s (2006) Distributed Situation Awareness theory. The theory states that SA is a property of systems that include many human and non-human agents. This perspective was inspired by Hutchins' (1995) work on distributed cognition evident in aircraft cockpits, which posits that non-human agents can possess cognitive attributes and that cognition at times depends on interactions with external tools and technologies. The distributed approach to SA is designed to capture the flow of information between human and non-human agents with the use of networks. It is an apt approach for assessing SA in widely distributed systems.

Another perspective on SA is the Situated SA approach (Chiappe et al., 2012). This approach differs from Distributed SA by stressing the interactions between individual operators and their immediate task environment, instead of focusing on entire socio-technical systems. Situated SA is also agnostic about whether entire systems can possess cognitive properties over and above those possessed by individual operators (Chiappe, Rorie, Morgan, & Vu, 2014). It is also more appropriate when considering individual operators, such as console operators.

The Situated Approach is based on the notion that human agents offload information to the environment to reduce the amount of cognitive work that must be done internally. In this way, the environment itself does not possess the awareness of that offloaded information, but instead retains the information until the human agent wishes to retrieve it (Chiappe et al., 2012). Contrary to Endsley's Three-Level Model, the situated approach holds that only some information is held in the mind instead of detailed situational models. This is consistent with perceptual research showing that the internal representations of a scene are incomplete (Clark, 2008; Rensink, 2000). Specifically, it has been noted that humans fail to notice large changes in the scenes they are viewing (Simons & Rensink, 2005). If large scale changes are not detected then perhaps humans do not generate accurate and detailed representations of their environment as Endsley (1995b) presupposes. In a Situated SA view, the storage of information may not involve storing all information internally, but instead only storing general cues on where to access more specific elements of information (Chiappe et al., 2012).

Additionally, this view holds that the process-product distinction is arbitrary, as part of operators' SA includes knowledge of the processes by which to access offloaded information (Chiappe, Strybel, & Vu, 2015). However, Chiappe et al. (2015) do not disregard an internalist view of SA entirely. Instead, the Situated approach argues that there are two kinds of information making up our SA. One includes information that is held internally, and the other is information that is offloaded, but readily accessible to the operator. Humans perceive information, understand it, and predict future situations with the aid of minimal internal representations and reiterative interactions with their environment (Chiappe et al., 2015). The information that is offloaded is specific in nature, as it has been found that operators such as air traffic controllers are more affected by an inability to access information when they are asked to answer questions about specific information as opposed to general information (Morgan, Chiappe, Kraut, Strybel, & Vu, 2012).

To measure Situated SA, Durso and Dattel's (2004) Situation Present Assessment Method (SPAM) can be used (Chiappe, et al., 2012; Chiappe, Vu, & Strybel, 2015). SPAM is a real-time probe technique. Although very similar to SAGAT, it differs in very important aspects. Operators are still queried as to their situation; however, unlike SAGAT, the displays are not blanked and the task is ongoing. Furthermore, SPAM queries are administered one at a time, whereas SAGAT prompts are administered as a battery of questions.

SPAM also incorporates "ready prompts," which warn operators that a question is in the queue. They are supposed to indicate when they are ready to receive the question. The use of a ready prompt means that this technique produces data other than just accuracies to the probes. Two separate response times are also produced; one from the onset of the ready prompt to the time the participant accepts the prompt (ready time), and one from the time the question is presented to the time that it is answered (response time). Since participants are instructed not to accept a ready prompt until their workload allows them to, the former response time is regarded as a measure of workload (Durso & Dattel, 2004). The latter response time is said to give a measure of SA. The response times to answering a probe question also provide an indication of whether or not information has been offloaded onto the environment (Chiappe et al., 2015). Specifically, if operators respond quickly to the question, then it is said that that type of information had been stored internally. If operators took a longer time to respond to the question, then the operators had to search their environment for the offloaded information.

In addition to providing more measurements (i.e., RT and accuracy) than SAGAT, SPAM does not require the task to be paused, therefore making it more readily applicable in testing environments other than simulations. SAGAT, on the other hand, is limited in its application to only simulated environments where the researcher has complete control over the situation.

Selection of Situation Awareness Measurement Tools

When choosing metrics for SA for domains such as petrochemical refining, an important factor to consider is a metric's psychometric properties. A measure of SA should have reliability and validity (Salmon, Stanton, Walker, & Green, 2006; Strybel et al., 2010). It should also have sensitivity, diagnosticity, and strong usability and user acceptance (Marras & Karwowski, 2010; Strybel et al., 2010; Strybel et al., 2011). "Reliability" refers to how consistently a metric captures the intended construct. "Validity" is the degree to which the measurement technique is actually measuring the intended construct. "Sensitivity" refers to the ability of the measurement technique to capture changes in the construct that are caused by a manipulation. "Diagnosticity" refers to the degree to which the measurement technique can identify factors that yield changes in the construct. "Usability" refers to the ease of using and implementing the measurement technique by the researcher. Finally, "user acceptance" refers to opinions about the measurement technique as well whether the measurement technique affects an operator's performance.

Ideally, a measurement technique would have strong results on each of these criteria. However, it is difficult, and often experimentally unfeasible to test for all in one study. This fact is exacerbated when one considers the plethora of ways that one can go about measuring SA. As we have seen, however, two very promising techniques for measuring SA are SAGAT and SPAM. Salmon et al. (2006) categorized these two measurement techniques as freeze/offline probes, and real-time/online probes, respectively. Probe methods are regarded as being the most objective and having high face validity as compared to other techniques such as subjective ratings (Strybel et al., 2011). SAGAT has been extensively used in several domains and has received considerable evidence as to its merit on objective measurement criteria (Endsley, 2012). SPAM, although not as applied in as many settings (Chiappe et al., 2015), has also received testing for its psychometric properties (Loft et al., 2014; Loft, Morrell, & Huf, 2013).

Validity

Much research has been done to validate probe techniques, as well as to determine whether online and offline methods differ in this regard (e.g., Durso et al., 1999; Durso et al., 2006; Endsley, 1995a). In examining probe measures' validity, Durso et al. (1999) conducted an experiment in an air traffic control context to examine whether SA measures add any predictive validity over and above mental workload. Participating air traffic controllers completed five separate 30 minute scenarios where they managed air traffic consisting of commercial and military aircraft. SA was measured with SPAM, SAGAT, and a subjective measure of SA known as the Situation Awareness Rating Technique (SART; Taylor, 1990) on three of the five scenarios. Workload was also measured for each of the scenarios. Performance was measured by Subject Matter Experts' (SMEs) subjective evaluations of the controllers and by remaining action counts, which refers to the remaining control actions not taken by the participants to move an aircraft out of the controller's airspace sector correctly. Regression analyses with workload and each of the three SA metrics were carried out to identify the best predictors of performance. Overall, all three SA techniques were shown to predict SME evaluations of performance. However, only SAGAT and SPAM significantly predicted the variance in the number of remaining controller actions.

In a more recent study, Durso et al. (2006) further investigated the validity of SA measures independently of other cognitive constructs. Given that SA relies on but is distinct from cognitive factors such as attention and working memory (Endsley, 1995b), the question posed was whether these SA metrics possess additional predictive validity after accounting for a battery of cognitive variables. The researchers asked 89 college-age participants to complete cognitive tests that included measures of general fluid ability, general crystalized ability, short-term memory, working memory, spatial memory span, closure flexibility, as well as some personality and demographic variables. In addition to the cognitive tests, participants were also tested on an air traffic control task, the Federal Aviation Administration's Air Traffic Scenarios Test (ATST).

In the study, SA was measured with SPAM and a SAGAT-like offline measure in six separate scenarios (Durso et al., 2006). Performance of the participants was measured using the ATST and included handoff delay time, which is the time from when an aircraft appears on the control screen to the time the participant clicks on the aircraft (which indicates the participant has begun control over the aircraft); the number of air traffic control errors, which include landing aircraft at incorrect altitudes, having aircraft collide, or having aircraft hit boundaries; and en route time, which is the time from when the participant assumes control of an individual aircraft until the aircraft disappears from their control screen.

SPAM was found to account for additional variance in handoff delay times above fluid intelligence whereas SAGAT was not (Durso et al., 2006). SPAM scores were also able to predict an additional 15% of the variance in air traffic control errors beyond that accounted for by closure flexibility, conscientiousness, and spatial working memory whereas SAGAT did not. In general, then, SPAM featured greater predictive validity. The only exception was in the case of en route times, where SAGAT featured greater predictive validity than SPAM. The study also found that SPAM and SAGAT did not intrude on operator performance, as performance was comparable in scenarios where SA was measured, and when it was not.

In a study that further sought to validate SPAM, Strybel et al. (2010) conducted an experiment in which eight pilots flew desktop-simulated aircraft on an arrival path into an airport that required appropriate spacing over the course of 12 trials. The trials varied on plausible future conceptualizations of operations in air traffic management and were designed to test spacing and conflict resolution responsibility allocation between pilots, air traffic controllers, and automation. Real-time SA probe questions that asked pilots about conflicts, communications, and aircraft status were presented to the pilots throughout the trials. Their results indicated that ready latencies were correlated with subjective workload ratings and uncorrelated with probe response latencies, suggesting that the ready latencies are indicative of participant workload. It was also found that SPAM probe categories (e.g., conflict or communication probes) were significantly correlated with the pilots' conflict resolution, merging, and spacing performance. SPAM probe response latencies in certain categories were also correlated with spacing performance while probe accuracy was not, suggesting that latencies may provide additional diagnostic information that accuracy does not.

Sensitivity

It is very important for a measure of SA to be sensitive to changes in SA that may be caused by factors such as display designs, automation, workload, or training programs. For example, in an investigation of SAGAT, Endsley (1995a) explored how the SA information provided by SAGAT can be used to improve display designs. Six pilots completed five trials which consisted of five different display configurations. The display configurations were manipulated with the goal of improving pilots' perceptions of target aircraft altitude. The results indicated that certain display configuration hurt the SA of the pilots regarding target range, heading, and azimuth, while SA regarding altitude was not affected.

Endsley, Sollenberger, and Stein (2000) ran a study to investigate the sensitivity of SAGAT and a SPAM-like online technique. Ten air traffic controllers performed a high-fidelity simulation in which they controlled air traffic with a traditional radar display or an enhanced radar display. Their results indicate that SAGAT was able to detect changes in SA between each of the display configurations whereas the online technique was not.

Nonetheless, a study by Vu et al. (2009) did find SPAM to be sensitive to changes in SA. Specifically, SPAM was found to detect differences in SA between student and experienced air traffic controllers. Experienced air traffic controllers were able to more accurately answer SPAM probes regarding the future whereas student air traffic controllers were not as apt at answering these types of questions. In a separate investigation, Vu et al. (2012) varied the traffic density in a simulation completed by air traffic control students. The air traffic control students completed a training program where they were trained to rely either on conventional air traffic management techniques or to rely on automated tools. SPAM probe questions were administered across these scenarios and it was found that SPAM was able to detect changes in SA caused by this manipulation of workload. Additionally, SPAM probes were found to be sensitive to changes in SA caused by the training manipulation. In short, both SAGAT and SPAM have shown evidence suggesting that they are sensitive to changes in SA in certain contexts.

<u>Intrusiveness</u>

Another important criterion to take into consideration when choosing a measurement technique is its intrusiveness to the primary task. Researchers such as Salmon et al. (2006) state that probe measures are by definition intrusive because the probes presented will always be a secondary task that draws attention from the primary task. However, proponents of SPAM and SAGAT maintain that the additional load provided by the administration of probe questions may not be great enough to cause changes in the primary task (Durso & Dattel, 2004; Endsley, 1995a).

Once again, much attention has been given to this issue for both SAGAT and SPAM (Bacon & Strybel, 2013; Endsley, 1995b; Silva et al., 2013). For example, Endsley (1995a) had participants complete trials in which simulation pause frequency and duration were manipulated on three levels each (1, 2, or 3 simulation pauses during a trial and 0.5, 1, or 2 minute pause durations). In addition, trials where no simulation pauses occurred were also presented to the participants. The results indicated that performance did not differ between trials where pauses occurred and when they did not. Overall, the evidence supported that SAGAT administrations were not intrusive to the primary task.

In a further attempt to examine the intrusiveness of SPAM, Bacon and Strybel (2013) investigated whether asking questions about the present situation can change the awareness of the operators. Offline and online SA techniques may change the SA of operators by giving operators cues as to what information is important. This could direct the attention of an operator to information that can bias the performance of the operator after a question. To test this, Bacon and Strybel manipulated the presentation of SPAM probe questions to coincide with certain events in the scenarios that air traffic control students completed. Their results show no evidence that SPAM probes altered the awareness or performance of the student controllers post-probe.

Silva et al. (2013) also examined potential intrusive effects of SPAM in an air traffic control context. Fourteen air traffic control students participated in two scenarios. The participants were split into groups denoting their proficiency in managing air traffic (whether they had achieved Journeyman status or not) and also according to whether their training emphasized the use of conventional air traffic management techniques or proposed automated tools. SPAM probes were administered in only one of the scenarios, thereby allowing the investigators a baseline comparison. Despite finding effects of journeyman status and effects of type of training, no evidence was found that SPAM probes had an effect on the air traffic control students' objective performance or perceived workload. This suggests that SPAM may not be intrusive to the primary task even though it is carried out at the same time.

To summarize, both probe techniques have received some support for their possession of desirable psychometric properties. As a result, they both warrant being considered as potential SA metrics for use in the petrochemical refining field.

Present Investigation

The petrochemical refining industry is seeking ways to optimize its processes in terms of safety and efficiency, while simultaneously reducing the environmental impact. One way to assist in this goal is to measure SA of operators to develop support systems that will improve performance. As we have seen, two prominent techniques are SAGAT and SPAM. The current investigation examined these two measures in terms of their psychometric properties with the goal of gathering evidence for the most appropriate technique for measuring SA of console operators in a petrochemical refining context. Currently, there is no universally accepted method for measuring SA, nor is there one derived from the petrochemical field. By having compared these techniques with the results of this investigation, it is be possible to identify if one of these measures is more sensitive, valid, and less intrusive.

Hypotheses and Predictions

The current study tested the two techniques on three measurement criteria previously identified as important (Salmon et al., 2006; Salmon et al., 2009; Strybel et al., 2010; Strybel et al., 2011). Since both techniques have received considerable attention and favorable results on their own merit, it is not hypothesized that one would be more effective than the other. Instead, it is argued that the most appropriate technique to use in petrochemical refining should be the most valid, sensitive, and least intrusive to console operator primary tasks.

This project tested SAGAT and SPAM on criterion related validity by associating the measures of SA produced by each measure with console operator performance. It is predicted that the most valid measure of SA would elicit stronger relationships with performance. Additionally, for SPAM, the latencies to accept a probe question should also be associated with another measure of operator workload, which in this investigation was be the validated NASA Task Load Index (NASA TLX; Hart & Staveland, 1988).

As mentioned above, neither Endsley (1995a) nor Silva et al. (2013) found evidence that either measure negatively affected performance or workload. To see if this is the same case in the petrochemical field, each technique was also tested for intrusiveness. This was done by comparing scenarios in which the techniques are individually administered with scenarios in which no technique was administered as a baseline. Performance and workload variables were assessed to capture whether the administration of either or both of these techniques affected console operator workload, performance, or both. The least intrusive technique would have no impact on either the performance on the primary task or on the perceived workload of the console operators.

Lastly, this project tested the techniques for their sensitivity to detect changes in console operator SA. This was done by manipulating the workload of the task (Roberts et al., 2012; Vu et al., 2009; Vu et al., 2012). Each technique, if sensitive, should then be able to detect changes in operator SA from a lower workload task to a higher workload task. The accuracies of each, and the response latencies of SPAM, should vary according to the task workload manipulation.

To summarize, the present study investigated how each measurement technique, SAGAT and SPAM, fares in terms of the measurement criteria of validity, sensitivity, and intrusiveness to informatively choose the most effect and appropriate one to use in future studies in the petrochemical field.

CHAPTER 2

METHODS

Participants

Participants were recruited from a pool of 13 current console operator incumbents at a large petrochemical refinery. A total of 11 console operators participated in this study after giving their informed consent (see Appendix A). These console operators manage the process of a hydrocracking plant that uses hydrogen to break down substances in crude oil. The incumbent pool has a wide range of experience; some operators with over a decade of experience, and some with only a few years. All received the same training when first hired, which ranged from basic training in outside field operations to training on the operations control console, all within the same processing plant. The participants were compensated with overtime pay.

Materials

Simulator and Scenario

A medium fidelity process control simulator modeled after the hydrocracking unit that the potential console operators currently work on was used. The simulation set up consists of four computers; three computers process the control console data and give video feed to two monitors each for a total of six monitors, and one computer processes the simulator controls which include scenario selection, start, pause, and stop (Figure 1). A fifth station communicates video feed from an adjacent room which provides the capability to simulate communications with outside field operators if the scenario requires it. Operators have access to all plant parameters through the use of any keyboard or display. That is, operators have the ability to pull up any information about the plant on any one of their six displays with any one of the three keyboards, just as in the real plant. The simulator collects numerous process variables as the simulation is ongoing with a 1 minute sampling rate.

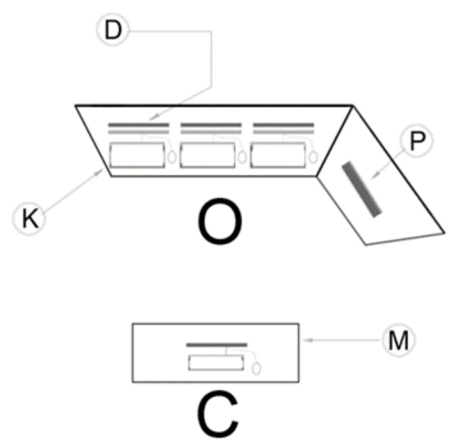

FIGURE 1. Overhead view of simulation setup. O is the operating participant, C is the confederate administrator, M is the Master control station, D is a display, K is a keyboard, P is the probe station.

The scenario chosen for this study is that of a loss of hydrogen feed to the hydrocracking plant. As stated above, the simulated plant uses large amounts of hydrogen to break down substances in crude oil. Quite frequently, the plant that supplies the hydrogen to the hydrocracker trips, resulting in a complete loss of hydrogen feed. The console operators must then respond to this loss of hydrogen by cutting down on processes within the hydrocracking unit in order to prevent damage to equipment and maintain high product quality specifications. This scenario was chosen in consultation with SMEs and company managers because of its frequent nature, difficulty, and monetary costs involved in managing the incident. It is possible to resolve this situation in 30 minutes, but can last for hours if the inappropriate actions are taken. For the purposes of this experiment, trials were limited to 33 minutes to balance data collection efforts and costs.

Subjective Measures

Three subjective questionnaire measures were collected from participants: the SART, NASA TLX, and a SA usability questionnaire. The SART is a measure of subjective SA based on three dimensions: attentional supply, demand, and understanding of the situation (Taylor, 1990). The *Understanding* dimension measures the quantity and quality of information available to the operator. *Demand* measures the instability, complexity, and variability of the situation. *Supply* measures the degree of arousal, concentration, and mount of free cognitive capacity, as well as the amount of divisions in attention that the situation requires. It contains 10 items with rating scales ranging from one to seven that are calculated to produce the three dimensions. A combined SART score can then be obtained by using the equation: SART Combine Score = Mean Understanding Rating – (Mean Demand Rating – Mean Supply Rating). To measure overall SA, a combined score was in this study.

The NASA TLX measures workload on six dimensions: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration (Hart & Staveland, 1988). The ratings on these dimensions can be added and multiplied by 1.11 to yield a combined score on a 100 point scale. Workload is considered a complex construct, and combing these dimensions allows researchers to capture operator workload holistically. The NASA TLX is the most commonly used subjective measure of workload, and has been shown to have acceptable levels of reliability and validity (Gawron, 2000).

The SA usability scale consisted of six questions that asked the operators about their experiences with answering questions with both techniques (refer to Appendix B for the full questionnaire). Participants were asked about how simple it was to answer questions during their task with both techniques on a 7-point Likert scale ranging from 1 “Very Simple” to 7 “Very Difficult.” Participants were also asked if they believed that answering questions during their task affected their ability to manage the scenario. Next, operators were asked to what degree their ability to manage the scenario was affected by answering questions with each technique, also on a 7-point scale ranging from 1 “Did not affect my ability at all” to 7 “Greatly affected my ability.” Lastly, operators were asked which method they preferred overall.

SPAM and SAGAT

Both SAGAT and SPAM queries were presented using a small touchscreen computer placed adjacent to the operators’ control board. Each technique was followed as true to their conceptions as possible, with the only modifications done to improve the usability of each technique. This was done in order to examine each of these measures globally in relation to the psychometric criteria that are being put forth in this project. Follow up studies will be conducted to examine the specific features of each technique that contribute to any changes that have arisen from the current project.

Recommendations for administering SPAM put forth by Strybel et al. (2011) were followed and were also applied to SAGAT as best possible. For SPAM, an audible tone sounded when a question was ready to be accepted on the touchscreen computer. After the operator accepted the question (for which they were instructed to only accept if their workload allows), the SA probe was presented. A total of nine SA probe queries were presented to the operators per trial; one about every 3 minutes and starting at 6 minutes into each scenario (Strybel et al., 2011). Responses to the queries were collected by having operators select their answer choice via pressing a button on the touchscreen. With SAGAT, probe questions were administered in a battery of three queries at three points in the scenario while the screens were blanked: First at about 7.5 minutes, second at about 15 minutes, and lastly at about the 22.5 minute mark. The numbers were not exact for either measure to make the expectation of a question unpredictable to the operator (Endsley, 1995a; Strybel et al., 2011). To keep the number of queries equal across both techniques, there were fewer presentations of SAGAT than there were of SPAM due to questions being presented in a battery instead of individually. SAGAT queries were also presented on the touchscreen computer and collected by touch responses on the screen. Situation awareness accuracies and both ready and response latencies for SPAM, were collected by the touchscreen computer.

Probe Queries

Probe queries were created with guidance from Hogg, Folleso, Strand-Volden, and Torralba (1995) in terms of information categories. Hogg et al. developed probe questions for a SAGAT-like measure of SA to be used in nuclear process control. They developed categories of information that relate important process parameters with time. Three categories are defined as queries relating the recent past to the present, the present state to the normal state, and the present state to the near future (Hogg et al., 1995). A total of 36 questions were constructed to be presented nine times in each of four probe-presented trials (see Appendix C). Guidance on the responses to the queries was also taken from Durso and Dattel (2004) and Morgan et al. (2012) to simplify the process. Specifically, responses to probe queries were limited to "Yes" and "No."

Experimental Design

The experimental design used is a 2 (Task workload: less workload and more workload) × 3 (situation awareness measurement: SPAM, SAGAT, and baseline). Both factors are repeated measures variables (see Table 1).

TABLE 1. Experimental Design

	SPAM	SAGAT	Baseline
Lower Workload	SPAM/Lower Workload	SAGAT/Lower Workload	Baseline/Lower Workload
Higher Workload	SPAM/Higher Workload	SAGAT/Higher Workload	Baseline/Higher Workload

Independent Variables

Scenario difficulty was manipulated in accordance with SME recommendations. The hydrocracking plant is a complex system, composed of several reactors, pumps, compressors, and distillation columns that interact not only with each other, but with other plants in the refinery. Two levels of difficulty were manipulated in such a way that the task becomes more difficult, but does not change the nature of the task. This was done by increasing or decreasing certain feed rates to and within the plant to make the scenario more difficult, but maintain the same behaviors to handle the situation. This is analogous to increasing the traffic density in air traffic control studies. Within each level of difficulty, three different scenarios were constructed. This was done by switching the loads on certain compressors and pumps. For example, if Compressor A was working at a 100% load, and Compressor B was at 80% load, then the compressor loading would be switched to still maintain a 90% average loading, but on different compressors. This ensured that identical behaviors were not taken on each scenario, yet all operators still performed the same task.

Within each level of difficulty, SAGAT, SPAM, and a no SA measurement technique scenario were experienced by the console operators, to yield a total of six trials. SPAM was administered on two trials, SAGAT on another two, and the baseline, no-probe on the final two. Situation awareness measurement administration, as explained above, was done through a touchscreen computer.

Dependent Variables

Several dependent variables were collected in this study. These include all process variables, which are indicative of operator performance in managing the plant during this incident, SA variables, and workload variables.

Performance. This study examined several performance variables. These variables were identified using a task analysis of the proposed critical scenario and were judged by SMEs as critical process parameters indicative of console operator performance. All performance variables were ranked in terms of importance by SMEs. The top eight most critical performance variables were selected, four of which are presented in this investigation, which is part of a larger study. These performance variables are values that will change over time depending on operator inputs and conditions of the plant. These four parameters can be displayed in any of the six displays that operators have access to. All four of these performance variables were analyzed using numerical integration methodology. Numerical integration was used to analyze the degree to which these performance measures were out of acceptable bounds and by how long. This methodology is heavily influenced by sampling rate and total amount of samplings and relies on these values to calculate the area under or over a curve. Since these scenarios were fixed in time (33 minutes long) and the sampling rate was always every 60 seconds, it was possible to sum the length of the deviations from the boundary to get a value of the degree and duration of time outside of bounds without having to account for sampling rate or total number of samplings. Thus, for all four performance variables, low values are indicative of better performance with that parameter. Figure 2 shows a graphical representation of this method, which can be applied to all four of the performance variables. With this methodology, lower scores in all four performance variables are indicative of better performance.

The first dependent variable was the combined temperature indication from four large compressors. These compressors move hydrogen into the plant, and when a loss of hydrogen occurs, they start to run empty. Since these compressors are usually set to automatically maintain a pressure output, they begin to work harder in response to lower amounts of hydrogen. If the compressors are not managed, high temperatures and excessive wear occurs that could damage the compressors. The upper limit for the temperature in these compressors is 300° F.

A second performance variable identified is the temperature in a distillation column. Distillation columns separate thicker oil products from lighter oil products such as gasoline. The temperature in these columns must be carefully controlled to ensure that

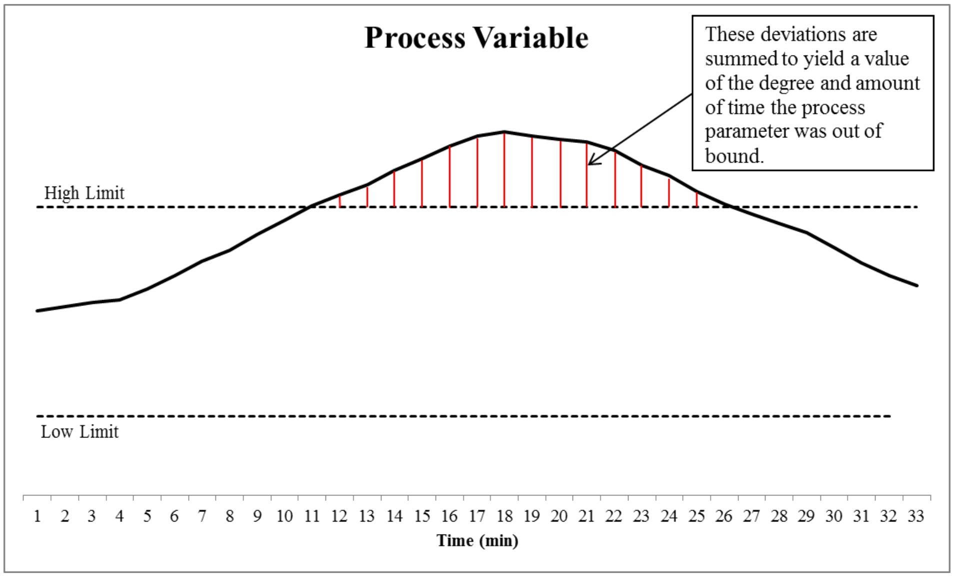

FIGURE 2. Example of numerical integration method for a given process variable.

products are distilled properly. In this incident, this specific distillation column temperature must be lowered because the fuel used to power the furnace that heats the column is diminished, but it must still be maintained above 605° F to ensure that the products inside are distilled properly and channeled to their next processes effectively. The upper limit for the temperature in this distillation column is 650° F.

The third performance variable refers to the calculated estimation of the temperature at which the jet fuel freezes. This is especially important for this plant and refinery, as it is a large supplier of jet fuel product to a large nearby airport. As a form of quality control, this value must be sustained during this incident. If the jet fuel freeze temperature is too high, the fuel would freeze inside the tanks of aircraft, which travel at high altitude and cold temperatures. A failure to maintain these specifications means that the products will have to be reprocessed, which is a high cost endeavor. The target temperature specification for jet fuel freeze will be -40° F.

The final performance variable is the calculated average bed (CAB) temperature in a reactor. Reactor CAB temperature is an indication of the temperature of each of the "beds" inside a reactor where chemical reactions take place. The temperatures in reactors are generally affected by how much material (feed) is being added to the reactor to cause those reactions. Since feed is typically cut in response to a low level of hydrogen, the temperatures in this reactor begin to fall, therefore not causing as much reaction as it should. If the materials in the reactors are not changed into the intended ones, the unchanged materials can harm reactions further down the process line. The CAB temperatures should be within ±3° F of the target value, which in this case is 750° F.

Situation Awareness. SA was measured in several different ways. SAGAT produced accuracies to probe questions administered in the respective scenarios. These were averaged to produce a mean of percent correct queries per participant in each scenario. Similarly with SPAM, the accuracies of each probe were averaged for each participant.

Situation awareness was also measured by the latencies to respond to correct SPAM probe questions (Durso & Dattel, 2004). The time between when the operator accepts a probe to the time the probe is answered correctly was averaged for all probes, each participant, and each condition.

Lastly, perceived SA was measured using SART (Taylor, 1990; see Appendix D). The questionnaire was administered after each trial, including the trials where SAGAT and SPAM were not administered.

Workload. Perceived workload was measured using the NASA TLX (see Appendix E). The TLX was also administered immediately after each trial, as well as after trials where SAGAT and SPAM were not administered. In SPAM, operators were instructed to not accept questions until their workload allowed. Therefore, workload was also measured using SPAM ready latencies.

Usability. The participants' thoughts and preferences regarding the method of answering questions (SPAM or SAGAT) during their task were also collected. This was done through a questionnaire that was administered after the study was completed and before they were dismissed (see Appendix B).

Procedure

Participants were recruited through word of mouth and postings near the plant console and plant field office. Participants were briefed on the procedure of the experiment and what was required of them. After informed consent, the participants were given information necessary to properly manage the plant. These data include overall refinery processing amounts, and specific product quality level benchmarks (these data typically change on a weekly basis and is decided upon by process engineers for refinery optimization).

Participants were then given three training blocks. The first training block lasted 20 minutes and was meant to familiarize the operators with operating the simulator, which is very similar to the consoles on which they work normally. The remaining two training blocks lasted 15 minutes each and were administered to introduce and practice the two SA techniques. These two training block were counterbalanced between participants in an ABAB pattern.

Six 30 minute experimental trials then followed the training blocks. As mentioned in the experimental design section, the six trials encompassed all combinations of task difficulty and probe technique. Subjective measures of SA and workload were administered after the completion of each trial for a total of six times (see Table 1). After the completion of the subjective measures, the participants were allowed to take a 10 minute break between all scenarios. After all experimental blocks were completed, participants were debriefed as to the purpose of the experiment and then dismissed. The complete study lasted roughly 8 hours, which was well within the range of a 12 hour shift that operators are accustomed to.

CHAPTER 3

RESULTS

Analyses were conducted on all the variables collected in this study, save for the SART measure, which is part of a different study. Separate analyses to address each of the goals of the experiment were conducted. An assessment of validity for each of the SA probe techniques was first conducted. Tests of sensitivity were also conducted on both techniques. An assessment of each probes technique's potential intrusiveness was examined as well as a subjective appraisal of each technique by the operators. See Table 2 for a complete list of performance variable means and variances.

Validity Assessment

To test for criterion validity of the SPAM and SAGAT techniques, Pearson's *r* correlation coefficients between measures of SA and the performance variables were calculated. Specifically, SAGAT and SPAM mean accuracies and SPAM mean response latencies were correlated with the means of each of the four performance variables with respect to each scenario. Regression analyses were conducted to examine if SA measures can predict performance above merely being associated wherever correlations were found to be statistically significant. It was predicted that the SA measures that most strongly predict performance scores have the strongest criterion related validity.

The analysis indicated that SPAM accuracies were significantly correlated with Distillation Column Temperatures in the low workload conditions, $r(11) = -.641$, $p =$

TABLE 2. Means and Variances for Performance Variables

Measure			Minimum	Maximum	Mean	SE
ComTemp						
	Low Workload					
		SPAM	0.29	92.19	22.14	7.70
		SAGAT	0.00	42.65	18.43	3.87
		Baseline	1.92	51.53	21.96	4.80
	High Workload					
		SPAM	3.84	37.35	17.91	3.59
		SAGAT	0.00	47.08	21.86	4.19
		Baseline	1.11	51.84	23.71	4.51
DisTemp						
	Low Workload					
		SPAM	0.00	33.18	7.59	3.51
		SAGAT	0.00	68.40	15.29	6.23
		Baseline	0.00	36.10	10.50	3.49
	High Workload					
		SPAM	0.00	105.59	39.11	11.09
		SAGAT	0.00	78.37	26.64	7.93
		Baseline	0.00	97.87	40.57	8.65
JetSpec						
	Low Workload					
		SPAM	0.00	151.94	26.52	13.75
		SAGAT	0.00	137.18	33.86	13.45
		Baseline	0.00	111.97	41.99	13.09
	High Workload					
		SPAM	2.87	196.36	44.47	16.70
		SAGAT	0.00	43.48	13.52	3.65
		Baseline	0.00	147.95	35.74	13.81
RxTemp						
	Low Workload					
		SPAM	160.94	1049.93	409.57	83.27
		SAGAT	156.81	1119.22	323.96	83.93
		Baseline	192.09	477.81	279.44	28.51
	High Workload					
		SPAM	214.01	1402.38	420.55	103.09
		SAGAT	173.59	608.30	388.78	42.42
		Baseline	207.67	675.71	369.35	46.07

Note. ComTemp = Compressor Temperature; DisTemp = Distillation Column Temperature; JetSpec = Jet Freeze Specification; RxTemp = Reactor Bed Temperature.

.034. Regression analysis followed this correlation to show that SPAM accuracies significantly predicted Distillation Column Temperature, $F(1, 9) = 6.262$, $p = .034$, $R^2 = .41$. The operator's performance lowered (i.e., improved) by -.641 standard units in temperature for each increased standard unit in SPAM accuracy, $\beta = -.641$, $t(9) = -2.502$, $p = .034$.

SPAM accuracies were also significantly correlated with performance on Jet Freeze Specification in the low workload scenario, $r(11) = -.653$, $p = .029$. SPAM accuracies additionally predicted a significant amount of variance in performance on Jet Freeze Specification, $F(1, 9) = 6.685$, $p = .029$, $R^2 = .43$. Operators performed -.653 standard units better on Jet Freeze Specification with each standard unit increase in SPAM accuracy, $\beta = -.653$, $t(9) = -2.586$, $p = .029$.

Lastly, SPAM accuracies showed an association with Reactor Bed Temperatures in the low workload scenario, $r(11) = -.688$, $p = .019$. Regression analysis found that SPAM accuracies accounted for 47.4% of the variance in performance with managing Reactor Bed Temperatures, $F(1, 9) = 8.107$, $p = .019$. Performance on Reactor Bed Temperatures improved -.688 standard units for each standard unit increase in SPAM accuracies, $t(9) = -2.847$, $p = .019$. A representation of all correlation tests can be found in Table 3.

Concurrent validity was assessed by performing Pearson's *r* correlation coefficients between mean SPAM ready latencies of all queries in the scenario and the mean of each of the operators' NASA TLX workload ratings for each scenario. It was predicted that if SPAM ready latencies are a measure of workload, then they should be able to predict the NASA TLX workload ratings for that scenario. However, no

TABLE 3. Complete List of Correlation Analysis

SA Measurement	Measure			
	ComTemp	DisTemp	JetSpec	RxTemp
Low Workload Scenarios				
SPAM Accuracy	-.456	-.641*	-.653*	-.688*
SPAM Response Latency	-.407	.034	.124	.135
SAGAT Accuracy	-.045	.334	.176	.485
High Workload Scenarios				
SPAM Accuracy	-.389	.098	.242	.344
SPAM Response Latency	.041	.464	.167	.139
SAGAT Accuracy	.468	.136	.195	.099

Note. ComTemp = Compressor Temperature; DisTemp = Distillation Column Temperature; JetSpec = Jet Freeze Specification; RxTemp = Reactor Bed Temperature.
*p < .05.

significant correlations were found between SPAM ready latencies and NASA TLX ratings in either low ($r(11) = .280$, $p = .405$) or high workload scenarios, $r(11) = .307$, $p = .358$, respectively.

Sensitivity Assessment

Sensitivity was tested using separate One-Way Repeated Measures Analyses of Variance (ANOVA) models with two levels of scenario difficulty for each SAGAT, SPAM, and Baseline condition. Before this, though, a check as to whether the manipulation of scenario difficulty worked was performed, aside from SME assurances. This was first examined with a two-way ANOVA with workload and SA measure administration as the independent variables and NASA TLX ratings as the dependent variable. Paired-sample *t*-tests were then conducted on performance between the less and more difficult scenarios for each SAGAT, SPAM, and Baseline conditions. Differences in workload and performance between low and high scenarios would indicate that the workload manipulation worked.

There was no overall main effect of workload ($F(1, 10) = .660$, $p = .435$) nor of SA measurement administration ($F(2, 20) = .800$, $p = .463$). There was, however, a marginally significant interaction between workload and SA measurement administration, $F(2, 20) = 2.762$, $p = .087$). Simple effects analysis revealed a significant difference in NASA TLX ratings between low and high workload conditions in the SAGAT scenarios, $t(10) = -2.32$, $p = .042$. Operators experienced greater workload in the high workload condition ($M = 60.36$, $SE = 3.92$) than in the low workload condition ($M = 47.90$, $SE = 5.01$). No other scenarios produced a difference in NASA-TLX ratings (see Table 4).

The workload manipulation caused a difference in performance on managing Distillation Column Temperatures in both Baseline, $t(10) = -2.951$, $p = .014$, and SPAM scenarios, $t(10) = -3.221$, $p = .009$. Within Baseline conditions, the workload manipulation caused better performance on managing Distillation Column Temperatures in the low workload scenario ($M = 10.50$, $SE = 4.80$) compared to the high workload scenario ($M = 40.57$, $SE = 8.65$). The same was found in SPAM conditions, where the low workload scenario caused better performance on managing Distillation Column Temperatures ($M = 7.59$, $SE = 3.51$) compared to the high workload scenarios ($M = 39.11$, $SE = 11.09$).

A difference was also found in managing Reactor Bed Temperatures between low and high workload conditions in the baseline conditions, $t(10) = -2.263$, $p = .047$. Performance in managing Reactor Bed Temperatures was worse in the high workload condition (*M* = 369.35, *SE* = 46.07) than in the low workload condition (*M* = 279.44, *SE* = 28.51. No other differences were found (see Table 4).

TABLE 4. Complete List of Manipulation Check Analyses

Measure	Low Workload		High Workload		*t*(10)	*p*	Cohen's *d*
	M	*SE*	*M*	*SE*			
Baseline Scenarios							
NASA TLX	54.13	4.03	52.86	4.66	.22	.829	.09
ComTemp	21.96	4.80	23.71	4.51	-.29	.779	.11
DisTemp	10.50	3.49	40.57	8.65	-2.95	.014	1.37
JetSpec	41.99	13.09	35.74	13.81	.49	.632	.14
RxTemp	279.44	28.51	369.35	46.07	-2.26	.047	.71
SAGAT Scenarios							
NASA TLX	47.90	5.01	60.36	3.92	-2.32	.042	.84
ComTemp	18.43	3.87	21.86	4.19	-.73	.484	.26
DisTemp	15.29	6.23	26.64	7.93	-1.10	.298	.48
JetSpec	33.86	13.45	13.52	3.65	1.44	.180	.62
RxTemp	323.96	83.93	388.78	42.42	-.93	.376	.29
SPAM Scenarios							
NASA TLX	57.41	5.33	55.47	4.92	.35	.731	.11
ComTemp	22.14	7.70	17.91	3.59	.71	.492	.21
DisTemp	7.59	3.51	39.11	11.09	-3.22	.009	1.16
JetSpec	26.52	13.75	44.47	16.70	-.80	.444	.35
RxTemp	409.57	83.27	420.55	103.09	-.07	.942	.04

Note. ComTemp = Compressor Temperature; DisTemp = Distillation Column Temperature; JetSpec = Jet Freeze Specification; RxTemp = Reactor Bed Temperature.

After this, tests for sensitivity were done by running the aforementioned three ANOVAs with task workload as the independent variable and SAGAT and SPAM accuracies and SPAM response latencies as the dependent measures. No SA measurement was found to be sensitive to changes in workload (see Table 5 for complete results).

TABLE 5. Complete List of Sensitivity Assessments

Variable	Low Workload		High Workload		$F(1,10)$	p	η^2
	M	*SE*	*M*	*SE*			
SAGAT Acc.	.68	.06	.67	.06	.01	.912	<.01
SPAM Acc.	.60	.07	.62	.07	.06	.817	.01
SPAM RespLat.	8450.04	724.32	9255.12	951.72	.81	.389	.08

Note. SPAM RespLate = SPAM Response Latencies.

Intrusiveness Assessment

Intrusiveness was analyzed by running separate One-Way Repeated Measures ANOVAs with three levels of measurement technique (SAGAT, SPAM, and baseline) on each of the four performance variables and the NASA TLX ratings separately for each workload condition. Measures of SA should not be intrusive to the operators' tasks in any way, nor should they change the way operators manage their tasks. It was expected that no effect of measurement technique will be found in any of these ANOVAs. However, a marginal effect of SA measuring technique was found for TLX ratings in the low workload scenarios, $F(2, 20) = 2.86$, $p = .081$, $\eta^2 = .22$. Post-hoc analysis showed that operators rated their workload to be lower during SAGAT scenarios ($M = 47.90$, $SE = 5.01$) as compared to the Baseline scenarios ($M = 54.13$, $SE = 4.03$; $p = .043$), and SPAM scenarios ($M = 57.41$, $SE = 5.33$, $p = .062$). There was no difference between SPAM and Baseline scenarios, $p = .493$. No other effects of SA measurement technique were found on any performance variable (see Table 6).

TABLE 6. Complete List of Intrusiveness Assessments

Measure	Baseline		SAGAT		SPAM		$F(2,20)$	p	η^2
	M	*SE*	*M*	*SE*	*M*	*SE*			
Low Workload Scenarios									
TLX	54.13	4.02	47.90	5.01	57.41	5.33	2.86	.081	.22
ComTemp	21.96	4.80	18.43	3.87	22.14	7.70	.21	.813	.02
DisTemp	10.50	3.49	15.29	6.23	7.59	3.51	.67	.521	.06
JetSpec	41.99	13.09	33.86	13.45	26.53	13.75	.90	.422	.08
RxTemp	279.44	28.51	323.96	83.94	409.57	83.28	1.04	.370	.10
High Workload Scenarios									
TLX	52.86	4.66	60.36	3.92	55.47	4.92	1.48	.253	.13
ComTemp	23.71	4.52	21.86	4.19	17.91	3.60	.94	.406	.09
DisTemp	40.57	8.65	26.64	7.93	39.11	11.09	.85	.441	.08
JetSpec	35.74	13.81	13.52	3.65	44.47	16.70	1.72	.205	.15
RxTemp	369.35	46.07	388.78	42.42	420.55	103.09	.17	.844	.02

Note. ComTemp = Compressor Temperature; DisTemp = Distillation Column Temperature; JetSpec = Jet Freeze Specification; RxTemp = Reactor Bed Temperature.

Usability Assessment

As an additional assessment, operators were asked to rate their subjective thoughts about each SA measurement technique on several questions after the study was completed. SPAM was referred to as "screens on and active" and SAGAT was referred to as "screens off and paused" in all questions. The first two questions asked, "How simple did you find answering questions while the screens were turned off and paused?" and "How simple did you find answering questions while the screens were on an active?" and were reported on a 7-point Likert scale ranging from 1 "Very Simple" to 7 "Very Difficult." On average, operators rated "screens turned off and paused" as 2.81 (*SD* = 1.25) and "on and active" as 3.36 (*SD* = 1.21) on the 7-point scale.

Another item asked the operators to answer the following question: "In general, do you feel that answering questions during your task affected your ability to manage the scenario?" with a binary "Yes/No" response. Six of the 11 (54.54%) operators answered

“No” to this question. Two following items consisted of “How much did answering questions while the screens were turned off and paused affect your ability to manage the scenario?” and “How much did answering questions while the screens were on and active affect your ability to manage the scenario?” and were also reported on a 7-point Likert scale ranging from 1 “Did not affect my ability at all” to 7 “Greatly affected my ability”. The average rating of operators for “turned off and paused” was 3.09 (*SD* = 1.14) and 3.55 (*SD* = 1.44) for “on and active.”

A final item asked the participants, “Overall, which method did you prefer to use?” with a dichotomous “Screens on and active/Screens off and paused” response. “Screens on and active” (SPAM) was preferred by the majority of operators (54.54%) for answering questions while completing their task.

CHAPTER 4

DISCUSSION

Situation awareness is an important construct that is highly related to operator performance. The petrochemical field is currently seeking ways to measure this construct to make effective decisions as to the design of console control board interfaces, operator training, and interactions with automation. Currently, there is no research on these two measures as they would be studied in the petrochemical field.

The methods that were used in this investigation are not new; they were recreated in this study to examine a set of tools in a new context that has not been looked at before. Both tools assessed here have been applied to a wide range of fields successfully and warranted consideration for use in petrochemical refining. This study investigated two widely used techniques for measuring SA: the Situation Awareness Global Assessment Technique and the Situation Present Assessment Method on the psychometric properties of validity, sensitivity, and intrusiveness. This was done by correlating the outputs of each SA technique with performance, testing the outputs for a change in SA, and testing a change in behavior potentially caused by administration of the technique.

The results for the validity assessment demonstrated that SPAM accuracies predicted performance on Distillation Column Temperatures, Jet Freeze Specification, and Reactor Bed Temperatures in low workload conditions only. This could be because operators may have generally been in a state of underload in low workload conditions and those operators that were aware in their task performed better than those that were not as aware. Thus, this would make this relationship more apparent in the low workload condition than in the high workload conditions. However, SPAM accuracies were not predictive of the fourth performance variable, Compressor Temperatures. It may be that these compressor temperatures, although rated as important for the outcome of the scenario, may not be prioritized by the operators. Lowering the priority of this variable would mean that it is not as related to awareness as the other performance variables. Another possibility is that the management of Compressor Temperatures is not indicative of human performance in this scenario. Lastly, from the values in Table 2, it is noted that variance was high across all performance variables. This may suggest that too much error variance was introduced by the way the operators managed their task during the study even though all participants received direction on managing the scenario and practice before the study began. The SMEs consulted for this study have mentioned that this was one of their own reasons for supporting this study; currently critical tasks are not performed in a standardized way as they should be. This and the lack of floor or ceiling effects suggest that error variance introduced by participants not only between them but also from trial to trial contributed to the lack of correlations found.

These results might be indicative that these performance variables that were associated with SA are related to the theoretical notions of the Situated Approach to SA. It could be that these performance variables are part of an integral feedback loop that operators create by interacting with their environment that otherwise was not captured by the SAGAT measure (Chiappe et al., 2015). A continual interaction with the environment--something that was not allowed by SAGAT--may have been needed to maintain an awareness that was predictive of performance (Chiappe et al., 2015).

Another finding in the validity assessment was that SPAM ready latencies were not associated with workload, indicating that they may not be a good measure of workload. However, NASA TLX ratings varied very little when compared across workload scenarios. In fact, NASA TLX ratings only varied from low to high workload conditions in SAGAT scenarios. These finding could be because the NASA TLX is a retrospective method administered at the end of trials whereas SPAM ready latencies were collected at several intervals throughout the trials. In this way, NASA TLX ratings may have been more affected by the happenings at the end of the scenario and SPAM ready latencies would be more robust against this effect. This would suggest that NASA TLX, in this study, may have not been the best benchmark to which to relate SPAM ready latencies to in the concurrent validity assessment.

In order to test for sensitivity of the SA techniques to changes in the SA construct, task workload was manipulated. Previous studies have shown that changes in workload can cause SA to change in different ways (Endsley, 1995b; Kaber & Endsley, 2004). Thus a manipulation check was conducted to determine that workload was indeed different between low and high task workload conditions. The workload manipulation was shown to be somewhat effective in certain cases. The data showed that operators rated their workload to be higher in the high workload condition in SAGAT scenarios. This pattern was also shown with performance, where the manipulation affected distillation column and reactor bed temperatures, but not the compressor temperature or the jet freeze specification. The fact that the manipulation did not have a wide-spread effect could likely mean that it was not a strong enough manipulation. Thus, this is likely the reason why no changes in any SA output were found in the data, indicating that neither measure was sensitive. It is unlikely that neither technique is insensitive especially since both techniques, and even more so SPAM response latencies, have been found to differ across workload conditions in previous studies (e.g., Endsley, 1995a; Endsley et al., 2000; Vu et al., 2009; Vu et al., 2012). Therefore, a likelier reason is that the workload manipulation employed was not strong enough to produce the intended effects on SA. Another important fact is that accuracies for both SPAM and SAGAT were very low. In the 60s percent range, they are not much higher than chance guessing (50%) since responses to probe queries were binary (see Table 5). The reasons for why accuracies are so low abound. It could be that operators did not understand the probe queries, despite having been validated with SMEs. It could also be that operators did not take seriously the SA probe task, deciding to prioritize performance on their primary task. Despite the reason, the fact that accuracies were near chance could also explain why the relationships that were expected were not found. If SAGAT and SPAM were being answered close to chance that could mean that operators were guessing, thus SA was not actually being captured and why the SA measures did not correlate with performance as intended.

The tests to determine if the administration of an SA measurement technique were intrusive were completed on perceived workload and performance. It was found that operators experienced marginally less workload when SAGAT was administered as compared to the baseline scenario in the low workload scenarios. Although the finding that workload was lower during SAGAT conditions compared to Baseline is at first glance favorable to SAGAT, it should be noted that SA measurement techniques should have *no* effect on workload. The fact that workload was lower in SAGAT conditions is evidence that SAGAT somehow changed the task for operators to make it seem easier. This could be due to how SAGAT removes the operator from the task, thereby giving them a small break from it while they answer questions that allows them recuperate before restarting their task. However, this was the only evidence found of an SA technique being intrusive. Neither SPAM nor SAGAT showed any other evidence of being intrusive upon operators in terms of workload or performance. Further, operators rated the difficulty of answering questions during their task between SPAM and SAGAT very similarly and relatively low overall, as shown in the post-study usability questionnaire. Operators also reported the intrusiveness of both techniques very similarly and, once again, relatively low. Lastly, a small majority favored using SPAM over SAGAT to answer questions during their task.

Overall, the results of this investigation found support for the use of SA measurement techniques in petrochemical refining. Specifically, the data showed an advantage in using SPAM. SPAM accuracies were found to significantly predict performance where no SAGAT accuracies did. Neither measure was sensitive to changes in workload, but it is likely that this was due to a relatively weak manipulation of task workload. Optimistically, the administration of SA techniques had very little or no effects on operator workload and performance. Here too lies an advantage for SPAM since the data showed an effect of SAGAT on perceived workload and SPAM also showed a slight preference from operators. In sum, the evidence shows that probe-type metrics such as SPAM, with refinement, can be used effectively in the petrochemical field to investigate issues related to SA.

Limitations and Future Directions

By testing these two techniques against three major measurement criteria, evidence for merits of use were found for SPAM. This could lead to further work on SA in this area, which should include the refinement of the SPAM tool to increase validity. One way that this could be done is by creating more effective probe queries that tap into the SA construct. A limitation of the current study is that it used recommendations for creating probe queries from the nuclear industry. Although seemingly analogous, it could be that the petrochemical field may need its own question scheme to be more valid in assessing SA.

Another limitation of this study is the low sample size. Efforts had to be taken so that the study was monetarily and operationally feasible to complete since actual operators were used. Although not unusual in applied work, efforts should still be made to obtain a large enough sample size. The nature of the task completed in this study was also not as realistic as it is in reality. If the operator were to experience this scenario on the actual console board, the event could realistically last hours. For feasibility, the scenario was cut short to about half hour blocks in order to efficiently complete the study. The simulator did, however, create the fidelity necessary to realistically manage the scenario as it would be done on the real plant.

The results of this investigation could help guide researchers and practitioners in choosing the SA measurement technique that would be most appropriate for the petrochemical field. Measuring SA could inform decisions regarding system, interface, and training changes. Understanding and describing the SA of console operators is the first step to improving operator SA by guiding the construction and development of training programs to improve their awareness. This has positive implications for the petrochemical refining industry. More situation-aware operators would detect, diagnose, and resolve incidents sooner. This would lead to less frequent and critical events, fewer environmental impacts, higher safety of workers, and more monetary gains instead of losses.

APPENDICES

APPENDIX A

INFORMED CONSENT FORM

CONSENT TO PARTICIPATE IN RESEARCH

How do Console Operators maintain their situation awareness?

You are being asked to participate in a study conducted by Tristan Grigoleit and Hector Silva from the Department of Psychology at California State University, Long Beach and the Learning and Development Department at Chevron, El Segundo as part of an ongoing development of new situation awareness metrics that are better suited for the improvement of training strategies. You were selected as a possible participant in this study because you are over the age of 18, are currently certified to work the Iso-side control console, have normal or corrected-to-normal vision, are not under the influence of any stimulant, medication, or other condition that may interfere with your ability to participate in this study, and are a current U.S. Citizen.

PURPOSE OF THE STUDY
We are examining how console operators maintain their awareness of what is going on as they manage the ISOMAX plant in a simulated environment, also known as their "situation awareness." Situation awareness refers to the ability to "have the picture," and be able to anticipate possible situations in one's plant.

PROCEDURES
If you agree to participate in this study, you will attend 50 minutes of training where you will familiarize yourself with the ISOMAX simulator and with the scenario you will be managing. Afterwards, you will participate in 3 hours of simulated scenarios over one day at the Refinery Optimization Center, Chevron, El Segundo. You will participate in six 33-minute scenarios that simulate a loss of hydrogen to the ISOMAX plant. After each scenario you will be asked to fill out questionnaires about your experience in the scenario. Rest breaks will be provided between scenarios. With rest breaks, training, scenarios, and lunch, participation in this experiment will take an estimated total of 8 hours.

POTENTIAL RISKS AND DISCOMFORTS
Because the tasks being performed are simulated, there is no higher risk than that of a typical day of computer work. The tasks are not being performed on the real console. Instead they are being performed in a safe and controlled environment. Breaks of appropriate length and frequency will be given throughout the day. A potential risk to participants is the breach of confidentiality of information collected during this experiment. To mitigate the effect of a breach of confidentiality, you will be assigned a number code that does not link any data collected from you in this experiment to your identity.

POTENTIAL BENEFITS TO SUBJECTS AND/OR TO SOCIETY
The results of this experiment will contribute to the development of situation awareness metrics better suited to the types of process control scenarios encountered in refining and to a better understanding of how operators such as yourself maintain SA.

PAYMENT FOR PARTICIPATION
You will receive overtime pay for each hour of participation in the study. Your hours of participation will be logged and your time card adjusted accordingly. Additionally, a gas card ($25 value) will be given for participation. Should you decide to withdraw from the

experiment before completion, compensation will be commensurate with your participation and you will still receive your complimentary gas card.

CONFIDENTIALITY

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law. The results from this study will not be associated with you in any way whatsoever.

PARTICIPATION AND WITHDRAWAL

You can choose whether to be in this study or not. If you agree to be in this study, you may withdraw at any time without consequences. Participation or nonparticipation will not affect your employment in any way. You will be paid for all sessions completed, as stated above. The investigator reserves the right to withdraw you from this research if he or she determines it is necessary to do so.

IDENTIFICATION OF INVESTIGATORS

If you have any questions or concerns about the research, please feel free to contact the Principal Investigators: Hector Silva (xxx-xxx-xxxx; hector.i.silva91@xxxxx.com and Tristan Grigoleit (xxx-xxx-xxxx; tgrigoleit@xxxxx.com) or the Faculty Advisor of this experiment Dr. Dan Chiappe (562-985-5024; dan.chiappe@csulb.edu).

RIGHTS OF RESEARCH SUBJECTS

You may withdraw your consent at any time and discontinue participation without any penalty. You are not waiving any legal claims, rights or remedies because of your participation in this research study. If you have questions regarding your rights as a research participant, contact the Office of University Research, CSU Long Beach, 1250 Bellflower Blvd., Long Beach, CA 90840; Telephone: (562) 985-5314 or email to ORSP-Compliance@csulb.edu.

SIGNATURE OF RESEARCH PARTICIPANT or LEGAL REPRESENTATIVE

I am at least 18 years old and I understand the procedures and conditions of my participation described above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been given a copy of this form.

____

Name of Subject

____ ______________

Signature of Subject Date

APPENDIX B

POST-STUDY QUESTIONNAIRE

P#:________

Post-Study Usability Questionnaire

1. How simple did you find answering questions **while the screens were turned off and paused** (circle one)?

1 -----------------2-----------------3-----------------4-----------------5-----------------6-----------------7
Very Simple ... Very Difficult

2. How simple did you find answering questions **while the screens were on and active** (circle one)?

1 -----------------2-----------------3-----------------4-----------------5-----------------6-----------------7
Very Simple ... Very Difficult

3. In general, do you feel that answering questions during your task affected your ability to manage the scenario (circle one)?

   Yes / No

4. How much did answering questions **while the screens were turned off and paused** affect your ability to manage the scenario (circle one)?

1 -----------------2-----------------3-----------------4-----------------5-----------------6-----------------7
Did not affect my ability at all ... Greatly affected my ability

5. How much did answering questions **while the screens were on and active** affect your ability to manage the scenario (circle one)?

1 -----------------2-----------------3-----------------4-----------------5-----------------6-----------------7
Did not affect my ability at all ... Greatly affected my ability

6. Overall, which method did you prefer to use (circle one)?

Screens on and active / Screens off and paused

Please explain why:____

APPENDIX C

PROBE QUERIES

PROBE QUERIES

1 In the last 3 minutes, did M1 DP (09PDC602) increase?
2 In the last 3 minutes, did Import Hydrogen from LSFO (09FI314) decrease?
3 In the last 3 minutes, did R1 Suction Pressure (09PC305A) increase?
4 In the last 3 minutes, did the 1000# Header (09PC305) decrease?
5 In the last 3 minutes, did the M2 Header (09PC503) increase?
6 In the last 3 minutes, did CKN Total Feed (05FI501) decrease?
7 In the last 3 minutes, did R620 CAB (06CC627) increase?
8 In the last 3 minutes, did R610 Feed Rate (06FC611) decrease?
9 In the last 3 minutes, did R530 CAB (05CC537) increase?
10 In the last 3 minutes, did R1 Suction Controller Valve Output (09PC305A.op) decrease?
11 In the last 3 minutes, did C730 Bottoms Level (07LC733B) increase?
12 In the last 3 minutes, did F720 Outlet Temperature (07TI720) decrease?
13 In comparison with steady state operation, is M1 DP (09PDC602) currently lower?
14 In comparison with steady state operation, is Import Hydrogen from LSFO (09FI314) currently higher?
15 In comparison with steady state operation, is R1 Suction Pressure (09PC305A) currently lower?
16 In comparison with steady state operation, is the 1000# Header (09PC305) currently higher?
17 In comparison with steady state operation, is the M2 Header (09PC503) currently lower?
18 In comparison with steady state operation, is CKN Total Feed (05FI501) currently higher?
19 In comparison with steady state operation, is R620 CAB (06CC627) currently lower?
20 In comparison with steady state operation, is R610 Feed Rate (06FC611) currently higher?
21 In comparison with steady state operation, is R530 CAB (05CC537) currently lower?
22 In comparison with steady state operation, is R1 Suction Controller Valve Output (09PC305A.op) currently higher?
23 In comparison with steady state operation, is C730 Bottoms Level (07LC733B) currently lower?
24 In comparison with steady state operation, is F720 Outlet Temperature (07TI720) currently higher?
25 In the next 3 minutes, will M1 DP (09PDC602) decrease?
26 In the next 3 minutes, will Import Hydrogen from LSFO (09FI314) increase?
27 In the next 3 minutes, will R1 Suction Pressure (09PC305A) decrease?
28 In the next 3 minutes, will the 1000# Header (09PC305) increase?
29 In the next 3 minutes, will the M2 Header (09PC503) decrease?

30 In the next 3 minutes, will CKN Total Feed (05FI501) increase?
31 In the next 3 minutes, will R620 CAB (06CC627) decrease?
32 In the next 3 minutes, will R610 Feed Rate (06FC611) increase?
33 In the next 3 minutes, will R530 CAB (05CC537) decrease?
34 In the next 3 minutes, will R1 Suction Controller Valve Output (09PC305A.op) increase?
35 In the next 3 minutes, will C730 Bottoms Level (07LC733B) decrease?
36 In the next 3 minutes, will F720 Outlet Temperature (07TI720) increase?

APPENDIX D

SITUATIONAL AWARENESS RATING TECHNIQUE

Participant: ______

Trial: ______

Scenario: ______

Situational Awareness Rating Technique (SART)

Please answer these questions with regard to the process situations presented in the scenario.

Instability of Situation
How changeable is the situation? Is the situation highly unstable and likely to change suddenly (high), or is it very stable and straight forward (low)?

1 Low	2	3	4	5	6	7 High

Complexity of Situation
How complicated is the situation? Is it complex with many interrelated components (high) or is it simple and straightforward (low)?

1 Low	2	3	4	5	6	7 High

Variability of Situation
How many variables are changing in the situation? Are there are large number of factors varying (high) or are there very few variables changing (low)?

1 Low	2	3	4	5	6	7 High

Arousal
How aroused are you in the situation? Are you alert and ready for activity (high) or do you have a low degree of alertness (low)?

1 Low	2	3	4	5	6	7 High

Concentration of Attention
How much are you concentrating on the situation? Are you bringing all your thoughts to bear (high) or is your attention elsewhere (low)?

1 Low	2	3	4	5	6	7 High

Participant: ______

Trial: ______

Scenario: ______

Division of Attention
How much is your attention divided in the situation? Are you concentrating on many aspects of the situation (high) or focused on only one (low)?

1 Low	2	3	4	5	6	7 High

Spare Mental Capacity
How much mental capacity do you have to spare in the situation? Do you have sufficient to attend to many variables (high) or nothing to spare at all (low)?

1 Low	2	3	4	5	6	7 High

Information Quantity
How much information have you gained about the situation? Have you received and understood a great deal of knowledge (high) or very little (low)?

1 Low	2	3	4	5	6	7 High

Information Quality
How good is the information you have gained about the situation? Is the knowledge communicated very useful (high) or is it a new situation (low)?

1 Low	2	3	4	5	6	7 High

Familiarity with Situation
How familiar are you with the situation? Do you have a great deal of relevant experience (high) or is it a new situation (low)?

1 Low	2	3	4	5	6	7 High

APPENDIX E

NASA TASK LOAD INDEX

Participant: ______

Trial: ______

Scenario: ______

NASA TLX Workload Scale

RATING SCALE DEFINITIONS

Title	Endpoints	Descriptions
MENTAL DEMAND	*Low/High*	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	*Low/High*	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	*Low/High*	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
EFFORT	*Low/High*	How hard did you have to work (mentally and physically) to accomplish your level of performance?
PERFORMANCE	*Good/Poor*	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
FRUSTRATION LEVEL	*Low/High*	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Participant: ______

Trial: ______

Scenario: ______

SAMPLE: NASA TLX RESPONSE FORM

MENTAL DEMAND

Low — High

PHYSICAL DEMAND

Low — High

TEMPORAL DEMAND

Low — High

PERFORMANCE

Good — Poor

EFFORT

Low — High

FRUSTRATION

Low — High

REFERENCES

REFERENCES

Bacon, L. P., & Strybel, T. Z. (2013). Assessment of the validity and intrusiveness of online-probe questions for situation awareness in a simulated air-traffic-management task with student air-traffic controllers. *Safety Science, 56,* 89-95.

Chiappe, D., Rorie, R. C., Morgan, C. A., & Vu, K. L. (2014). A situated approach to the acquisition of shared SA in team contexts. *Theoretical Issues In Ergonomics Science, 15*(1), 69-87. doi:10.1080/1463922X.2012.696739

Chiappe, D. L., Strybel, T. Z., & Vu, K. P. L. (2012). Mechanisms for the acquisition of situation awareness in situated agents. *Theoretical Issues in Ergonomics Science, 13*(6), 625-647.

Chiappe, D. L., Strybel, T. Z., & Vu, K. P. L. (2015). A situated approach to the understanding of dynamic situations. *Journal of Cognitive Engineering and Decision Making, 9*(1), 33-43.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension: embodiment, action, and cognitive extension*. New York, NY: Oxford University Press.

Cutchin, M. P., Martin, K. R., Owen, S. V., & Goodwin, J. S. (2008). Concern about petrochemical health risk before and after a refinery explosion. *Risk Analysis, 28*(3), 589-601.

Department of Industrial Relations Division of Occupational Safety and Health. (2014). *Process safety management for refineries* (Proposed general industrial safety order 5189.1, 2nd Draft). Retrieved from http://www.dir.ca.gov/dosh/DoshReg/Process-Safety-Management-for-Refineries/PSM-Draft-Regulation.2014-10-31.pdf

Durso, F. T., Bleckley, M. K., & Dattel, A. R. (2006). Does situation awareness add to the validity of cognitive tests? *Human Factors: The Journal of the Human Factors and Ergonomics Society, 48*(4), 721-733.

Durso, F. T., & Dattel, A. (2004). SPAM: The real-time assessment of SA. In S. Banbury & S. Trembley (Eds.), *A cognitive approach to situation awareness: Theory, measures and application* (pp. 137-154). New York, NY: Aldershot.

Durso, F. T., Hackworth, C. A., Truitt, T. R., Crutchfield, J., Nikolic, D., & Manning, C. A. (1999). Situation awareness as a predictor of performance in en route air traffic controllers. *Air Traffic Control Quarterly, 6*, 1-20.

Durso, F. T., & Sethumadhavan, A. (2008). Situation awareness: Understanding dynamic environments. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 442-448.

Endsley, M. R. (1995a). Measurement of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 65-84.

Endsley, M. R. (1995b). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32-64.

Endsley, M. R. (2012). *Designing for situation awareness: An approach to user-centered design*. New York, NY: CRC Press.

Endsley, M. R. & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, *42*(3), 462-492.

Endsley, M. R., Sollenberger, R., & Stein, E., (2000). Situation awareness: A comparison of measures. In *Proceedings of the human performance, situation awareness and automation: User-centered design for the new millennium* (pp. 15-19). Savannah, GA: SA Techniques.

Gawron, V. J. (2000). *Human performance measures handbook*. Mahwah, NJ: Lawrence Erlbaum.

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. 139-183). Amsterdam, The Netherlands: Elsevier.

Hogg, D. N., Folleso, K., Strand-Volden, F., & Torralba, B. (1995). Development of a situation awareness measure to evaluate advanced alarm systems in nuclear power plant control rooms. *Ergonomics*, *38*(11), 2394-2413.

Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science*, *19*(3), 265-288.

Kaber, D. B., & Endsley, M. R. (1997). Out of the loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Safety Progress*, *16*(3), 126-131.

Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science*, *5*(2), 113-153.

Lau, N., Jamieson, G. A., & Skraaning, G. (2013). Distinguishing three accounts of situation awareness based on their domains of origin. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 57, No. 1, pp. 220-224). Thousand Oaks, CA: SAGE.

Loft, S., Bowden, V., Braithwaite, J., Morrell, D. B., Huf, S., & Durso, F. T. (2015). Situation awareness measures for simulated submarine track management. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(2), 298-310.

Loft, S., Morrell, D. B., & Huf, S. (2013). Using the situation present assessment method to measure situation awareness in simulated submarine track management *International Journal of Human Factors and Ergonomics*, *2*(1), 33-48.

Marras, W. S., & Karwowski, W. (2010). *Fundamentals and assessment tools for occupational ergonomics*. New York, NY: CRC Press.

Morgan, C., Chiappe, D., Kraut, J., Strybel, T. Z., & Vu, K. P. L. (2012). An investigation into the potential sources of interference in SA probe techniques. In S. Landry (Ed.), *Advances in human aspects of aviation* (pp.623-631). Boca Raton, FL: CRC Press.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making*, *2*(2), 140-160.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, *7*(1-3), 17-42.

Roberts, Z., Silva, H., Vo, T., Boles, R., Gomes, N., Iwan, M., . . . Strybel, T. Z. (2012). Awareness of need for immediate communication predicts sitation awareness and performance in an ATC training course. In: Landry, S. J. (ed.) Advances in Human Aspects of Aviation, Ch. 20, HCII 2012, pp. 204-213. New York, NY: CRC Press.

Saleh, J. H., Haga, R. A., Favarò, F. M., & Bakolas, E. (2014). Texas City refinery accident: Case study in breakdown of defense-in-depth and violation of the safety–diagnosability principle in design. *Engineering Failure Analysis*, *36*, 121-133.

Salmon, P. M., Stanton, N. A., Walker, G., & Green, D. (2006). Situation awareness measurement: A review of applicability for C4i environments. *Applied Ergonomics*, *37*(2), 225-238.

Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D., Ladva, D., Rafferty, L., & Young, M. (2009). Measuring Situation Awareness in complex systems: Comparison of measures study. *International Journal of Industrial Ergonomics*, *39*(3), 490-500.

Silva, H. I., Ziccardi, J., Grigoleit, T., Battiste, V., Strybel, T. Z., & Vu, K. P. L. (2013). Are the intrusive effects of SPAM probes present when operators differ by skill level and training? In *Human interface and the management of information. Information and interaction design* (pp. 269-275). Berlin, Germany: Springer.

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, *9*(1), 16-20.

Stanton, N., Stewart, R., Harris, D., Houghton, R., Baber, C., McMaster, R. . . . Green, D. (2006). Distributed situation awareness in dynamic systems: Theoretical development and application of an ergonomics methodology. *Ergonomics*, *49*(12-13), 1288-1311.

Strybel, T. Z., Vu, K. L., Bacon, L. P., Kraut, J., Battiste, V., & Johnson, W. (2010, October). Diagnosticity of an online query technique for measuring pilot situation awareness in NextGen. In *IEEE/AIAA 29th Digital Avionics Systems Conference (DASC)* (pp. 4-B). Piscantaway, NJ: IEEE.

Strybel, T.Z., Vu, K.L., Chiappe, D., Battiste, V., Johnson, W., & Dao, Q.V. (2011) *Metrics for situation awareness, workload, and performance: Manual for online probe administration*. (Report No. NASA/TM-2011-215991). Mountain View, CA: NASA Ames.

Taylor, R. M. (1990). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In *Situational awareness in aerospace operations* (AGARD-CP-478; pp. 3/1-3/17). Neuilly Sur Seine, France: NATO - AGARD.

U.S. Chemical Safety and Hazard Investigation Board. (2007). *BP America refinery explosion and fire investigation report*. (Report No. 2005-04-I-TX). Washington, DC: Author.

U.S. Nuclear Regulatory Commission. (2012). *Human factors engineering program review model*. (Report No. NUREG-0711). Washington, DC: Author.

Vu, K. P. L., Minakata, K., Nguyen, J., Kraut, J., Raza, H., Battiste, V., & Strybel, T. Z. (2009). Situation awareness and performance of student versus experienced air traffic controllers. In *Human interface and the management of information. Information and interaction* (pp. 865-874). Berlin, Germany: Springer.

Vu, K. P. L., Silva, H., Ziccardi, J., Morgan, C. A., Morales, G., Grigoleit, T., . . . Battiste, V. (2012). How does reliance on automated tools during learning influence students' air traffic management skills when the tools fail? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 56, No. 1, pp. 16-20). Thousand Oaks, CA: Sage.