The Effect of Situation Presence Assessment Method (SPAM) on Air Traffic Control Students' Workload and Performance in High-Fidelity Simulations

by

Chao Zhang

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science in Technology

Approved April 2016 by the
Graduate Supervisory Committee:

Mary Niemczyk, Chair
Robert Nullmeyer
Michael Pearson

ARIZONA STATE UNIVERSITY

August 2016

ABSTRACT

This study examined the impact of Situation Presence Assessment Method (SPAM) administration on air traffic control (ATC) students' task workload and performance in high-fidelity ATC simulations. ATC students performed high-fidelity en-route simulations in two conditions: baseline conditions (without SPAM questions) and SPAM conditions. The data collected show that while workload in the two conditions were not significantly different, there was a trend of higher mental workload in SPAM conditions than in baseline conditions. Performance immediately following SPAM questions was revealed to be poorer than that preceding the SPAM questions and that over the equivalent time periods in the baseline conditions. The results suggest that a "Ready" signal before a SPAM question may not be enough to eliminate the impact of SPAM administration on ATC students' workload and performance in high-fidelity en-route simulations.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The study of situation awareness (SA) has a long history, and the concept of SA was first widely adopted by human factors scientists in the 1990s. It is a field of study closely related to complex, dynamic areas including air traffic control (ATC), power plant operations, and military command and control. While there is no universally accepted definition of SA, the most popular one was developed by Mica Endsley (1995b), who believes there are three levels of SA. Level 1 SA is the perception of relevant elements and their status in a situation. Level 2 is the comprehension and interpretation of these elements to achieve a general idea of what is currently happening. Level 3 is the projection of the future status of the elements to infer the future situation.

Human senses, cognitive abilities and knowledge of a specific task are required to develop good SA. For example, sight and hearing can assist in perception of a situation. Expertise, task experience, and cognitive abilities like attention, short-term memory, working memory, and long-term memory, underlie and influence perception, the understanding, and the projection of a situation (Endsley, 1995b). Good decision making usually is assumed to be associated with good SA. The Australia Department of Defense, Defense Science and Technology Organization (2005) conducted a study and found positive correlations between SA and decision making in reconnaissance simulations. Endsley (1990) found SA to be significantly related to subjects' performance, if they had expertise and experience at a task. This result indicates that SA can be considered as a positive factor of task performance.

To improve the quality of ATC education, colleges and ATC programs utilize simulators that can vividly simulate ATC scenarios to enable students to receive enhanced ATC training. For example, in the Arizona State University (ASU) ATC laboratory, the Adacel system is used to simulate a hybrid en-route ATC system, which is between the Automated Radar Terminal System (ARTS) and the Standard Terminal Automation Replacement System (STARS). Each ATC student's simulation performance is evaluated by former air traffic controllers utilizing an evaluation report. The report helps evaluate performance in a series of tasks such as airplane separation, coordination, traffic judgement, control methods, and radiotelephone communicaton. After observing a student's performance in the simulation, an experienced controller will comprehensively grade his/her performance. However, little information and feedback regarding a students' SA during an ATC simulation is available to the them. Without this information, students may not be able to know whether they have good SA during a simulation and how they can improve their SA. In aviation, SA is viewed as a crucial factor in operational safety. The Tenerife airport disaster was a runway collision between two Boeing 747s. It is one of the biggest aviation accidents in the history, in which 583 people were killed. Weick (1990) pointed out that the accident was caused partly by the controller's lack of SA. According to the Aviation Safety Reporting System (ASRS) database, there were 236 en-route controller reported flight incidents related to SA in 2015 in the U.S. Therefore, it is necessary to apply some SA measurement to ATC students' simulation training, which can help measure students' SA and provide extra information about how they can improve their ATC simulation performance.

The Situation Presence Assessment Method (SPAM) was chosen to be the SA assessment method in the current study. SPAM is an online-probe technique (Durso & Dattel, 2004), which means the simulation is not halted by the SPAM questions. The questions are asked several times in a simulation and can occur at any moment during the simulation process. Before a SPAM question is asked, a “Ready” signal is sent to the subjects. If the subject sends a “Yes” response to the “Ready” signal, the SPAM question is asked immediately afterward. Subjects in a simulation can also refuse the “Ready” signal until they feel ready for the SPAM question. The questions focus on subjects’ perception, understanding, and projection of the simulation situation. The accuracy of the answers and the response latency to the SPAM questions are collected to measure SA in the online-probe technique.

Stanton et al. (2013) state that a good measurement should be valid, reliable, sensitive, and diagnostic. Besides these characteristics, a good measurement should also not be intrusive. To be specific, the application of the SPAM technique should not increase the primary task workload. At the same time, ATC simulation performance should not be affected by the SPAM administration.

Durso and Dattel (2004) think a “Ready” signal before every SPAM question can successfully eliminate the impact of SPAM questions on task workload. However, compared to only performing a simulation, it is reasonable to think that responding to the “Ready” signals during a simulation may cause some disruption and add extra task workload to the subjects. When subjects answer the SPAM questions during a simulation, it can be considered dual-tasking. The simulation is the primary task, while answering the SPAM questions is the secondary task. Research by Pashler (1994) shows that dual-task is very likely to cause interference between the tasks performed at the same time. Driving a car and using a mobile phone simultaneously is a typical dual-task scenario. Strayer and Johnston (2001) conducted a similar experiment with experienced drivers and found a significant decrease in driving performance when the drivers were distracted by a mobile phone. Therefore, the similar interference may be expected in ATC simulations with the SPAM administration.

Several studies addressed this problem in Air Traffic Scenario Test (ATST) simulations (Durso, Bleckley, & Dattel, 2006; Pierce, Vu, Nguyen, & Strybel, 2008; Pierce, 2012). Durso et al. (2006) suggests that the administration of SPAM technique may not affect subjects' performance in the ATST simulations. However, Pierce et al. (2008) provides evidence that the application of SPAM may affect subjects' ATST task performance. Pierce (2012) also found that subjects' performance was affected by the administration of the SPAM technique. In Pierce's (2012) study, performance immediately preceeding and following the "Ready" signal was measured in the ATST scenarios and a decrease in performance immediately following the "Ready" signal was revealed, which provides evidence of the impact on performance caused by the SPAM administration in the ATST tasks. In these two studies conducted by Pierce et al. (2008) and Pierce (2012), the workload data shows that the administration of the SPAM technique in the ATST simulation didn't add extra task workload to the subjects.

Although these studies shed some light on the relationship between SPAM, workload, and performance, all of them used ATST, which is a low-fidelity ATC simulation, and chose subjects that had no ATC experience. No similar study has been done in high-fidelity simulations with subjects who do have ATC experience. Due to the difference between ATST and high-fidelity simulations, and the difference between subjects with and without ATC experience, it is unknown whether the relationship between SPAM, workload, and performance revealed in low-fidelity ATC simulation studies in which subjects are without ATC experience remains the same in high-fidelity ATC simulations with subjects that have ATC experience. Therefore, in the current study, the objective is to examine the effect of the SPAM technique on ATC students' workload and task performance in high-fidelity en-route ATC simulations.

To examine the effect, two conditions were designed in the study: baseline conditions and SPAM conditions. In baseline conditions, subjects participated in the simulations without the SPAM administration, while in SPAM conditions, subjects were required to participate in the simulations and answer SA related questions simultaneously. In the two conditions, several types of data were collected to indicate subjects' workload, and performance.

As no significant difference between workload data was found in the previous study (Durso et al. 2006; Pierce et al. 2008; Pierce, 2012), it is hypothesized that if SPAM is applied in high-fidelity en-route simulations, ATC students' workload would not be higher than that in baseline conditions. As dual-task is very likely to cause interference (Pashler, 1994) and previous studies (Pierce et al., 2008; Pierce, 2012) provide evidence that the administration of SPAM may influence primary task performance, it is hypothesized that if SPAM is applied, ATC students' performance in high-fidelity en-route simulations would be poorer than that in baseline conditions.

CHAPTER 2

LITERATURE REVIEW

Air Traffic Control

According to Nolan (1994), ATC is a service provided by air traffic controllers to pilots. The ATC service is typically divided into Tower, Terminal Radar Approach Control (TRACON), and en-route. It is the controllers' job to coordinate the movement of every airplane on the ground or in the airspace by communicating with pilots with radiotelephone equipment. The two primary role of the controller are to ensure flight safety and expedite air traffic flow. In the U.S. airspace, en-route controllers usually deal with flights that are at relatively high altitudes. En-route controllers' responsibilities include instructing pilots to climb or descend their aircraft to their assigned flight levels, ensuring the appropriate horizontal and vertical separation among all of the airplanes, and providing a variety of information such as navigation points and weather. To solve potential traffic conflicts and expedite air traffic flow in the airspace, en-route controllers can instruct pilots to change their aircraft status such as direction, speed and altitude. Because controllers need to know the current traffic situation and predict how the traffic will be in the near future, the en-route ATC position requires controllers to have very good SA at all times while on duty.

Situation Presence Assessment Method

The prototype of SPAM first appeared in the study conducted by Durso et al. (1995), which was called online queries with the situation present. In that study, chess experts, intermediate chess players, and novices were asked to monitor high-level chess games that were played by computers. When monitoring the games, subjects tried to anticipate the loss of chessmen with the help of a joystick. For example, pushing the stick indicated that a subject believed a chessman would be captured in the next several moves. In the study, subjects monitored the chess games and anticipated the loss of chessmen in four conditions: self-reporting conditions, eye-tracking conditions, situation-absent query conditions and situation-present query conditions. In self-reporting conditions, subjects were asked to discuss the chess games while monitoring the games and anticipating the loss of chessmen. Their discussion and comments were collected and transformed into quantitative data. In eye-tracking conditions, subjects were asked to wear eye-tracking equipment during the games and their eye-tracking data, such as fixations and saccades, was collected. In situation-absent query conditions, the chess games were halted, the chessmen on the chessboard disappeared, and questions about current and future chessmen positions were asked. In situation-present query conditions, subjects were asked similar questions while the games were not halted and the chessmen remained in view. In the situation-absent query and the situation-present query conditions, the data of the correct rate of answers and the latency of correct answers were collected. In all four conditions, the data of the time interval between their anticipation and when the loss of a chessman occurred was collected to determine subjects' SA. Durso et al. (1995) speculated that experts have better SA than novices. Therefore, chess experts were supposed to have the shortest time interval between their anticipation and when the loss of a chessman happened, while novices should have the longest time interval. The data of the time interval in the four conditions matched the assumption. However, among the four types of data collected in the four conditions (self-reporting data, eye-tracking data, the data of the correct rate and the latency of correct answers in situation-present query conditions and situation-absent query conditions), the data of the correct rate and the latency of correct answers in situation-present query conditions showed the most distinguishable differences among experts, intermediate players, and novices, which indicates that the situation-present query conditions can reflect the differences in SA better than the other three conditions. The data of the latency of correct answers in the situation-present query conditions was also better at differentiating among the three levels of chess players than the data of the correct rate of answers in the situation-absent query conditions. Based on the results, Durso et al. (1995) suggested that the best way to measure SA is to ask SA related questions while the situation remains in view to the subjects and to collect the data of the latency of correct answers.

Durso et al. (1998) developed the prototype of SPAM, online queries with situation present, and changed its name to SPAM. In their study, Durso et al. (1998) assumed that an air traffic controller's good SA enables him/her to know where to find the required information for performing an ATC task rather than to memorize some specific information. For example, a controller may not be able to remember the callsign of an airplane, but good SA helps the controller quickly know where to find this airplane on the radar screen when the controller hears the callsign. Jeannot, Kelly, and Thompson (2003) conducted structured interviews with air traffic controllers at the EUROCONTROL Institute of Air Navigation Services (IANS) to collect controllers' view of SA, in which the controllers' descriptions agreed with the assumption. Based on the assumption, Durso et al. (1998) redesigned the SPAM questions to make them be in favor of the way an controller remembers the traffic information. For example, instead of being asked what is the speed or heading of an airplane, a controller may be asked what is the relative location of one airplane towards another one on the radar screen right now.

Other Situation Awareness Measurement Techniques

Besides SPAM, researchers have developed a variety of methods to measure SA such as physiological techniques, subjective techniques, and probe techniques.

The P300 wave and eye-tracking techniques are good examples of physiological techniques. The P300 wave can reflect processes involved in stimulus evaluation (Polich, 2007). In scenarios where meaningless items are mixed with target items, the P300 wave will be elicited when a person reacts to a target item, while there is no occurrence of P300 wave when he/she reacts to the meaningless items. Therefore, the presence and timing of the P300 waves are usually used as an indicator of the perception of relevant information in a situation. However, it cannot reflect the other two levels of SA, which are the understanding and the projection of a situation. One good example of the eye-tracking technique is the NAC Eye-mark recorder. This device measures the point of gaze through the use of three cameras (Holmgvist, 2011). The data of the places where a subject looks can provide clues about how he/she perceives, comprehends the current situation, and infers the future status. But the fact that a subject looks at a point does not mean he/she collected the information. Therefore, the eye-tracking technique is unable to successfully reflect a subject's entire SA.

Several subjective techniques were developed such as self-report and self-rating. Self-report means to let the subjects describe the SA they had during the simulation. It is considered to provide a rich set of qualitative data, but it is hard to compare the data between two different self-reports. A typical example of self-rating is the Situation

Awareness Rating Technique (SART) (Taylor, 1989). SART uses 10 dimensions such as familiarity of the situation, attention, quality of information, quantity of information, and complexity of the situation, to measure a subject's SA. Subjects rate the dimensions on a 7-point scale after simulations or tasks. In the scales, 0 means low and 6 means high. Taylor and Selcon (1994) then improved SART and redesigned a simplified version, 3 Dimensions SART (3D SART). Instead of using 10 subscales, 3D SART measurement consists of three dimensions, which are supply of attentional resources, demands on attentional resources, and one's understanding of the situation. However, a big disadvantage of self-rating tools is that they are considered to be unable to reflect the actual level of SA when subjects are unaware that they missed some important information in the simulation. SART and 3D SART are post-trial techniques, which can successfully assess subjects' SA concerning the end of a simulation. However, due to working memory and short-term memory limitations, post-trial methods usually fail to perfectly reflect the SA in the beginning and middle parts of a simulation (Marois & Todd, 2004).

In general, probe techniques are the most widely studied and utilized SA measurements. There are two types of probe techniques. One is online and the most representative method is SPAM. The other one is offline and the most common example is the Situation Awareness Global Assessment Technique (SAGAT) (Endsley, 1995a). In SAGAT, when subjects are performing a simulation task, the simulation is halted several times. During the halts, the information on simulation displays disappears and questions related to their SA prior to the halts are asked. Due to time limitation, only a random selection of the SAGAT questions are asked during these time periods. The subjects' SA is measured by the rate of correct answers to the SAGAT questions.

Although online and offline probe techniques are both widely validated and applied, online probe techniques such as SPAM are considered to be better than SAGAT in some aspects when they are used to measure SA in ATC simulations. First, in SAGAT, the correct rate of answers to the SA related questions is used to indicate subjects' SA. In SPAM, besides the data of correct rate of answers, the response latency to the SA related questions is also collected to reflect subjects' SA. According to Paige and Strybel (2013), response latency is more sensitive than correct rate of answers when measuring subjects' SA. Second, in SAGAT, during the freeze time, there is no information on the screens. Chiappe, Strybel, and Vu (2012) state that the SA measured by SAGAT would be limited to the information stored in subjects' short-term memory and working memory. However, there is no such concern when SPAM is used because information on simulation displays does not disappear when subjects answer SA related questions. Third, in SAGAT, SA related questions require subjects to remember a specific piece of information such as the altitude of an airplane. Endsley and Rodgers (1998) found that sometimes controllers are very poor at answering SA related questions in SAGAT. On the contrary, SA related questions in SPAM are designed to be in favor of the way controllers store the information when performing an ATC task. Finally, SAGAT requires several halts during a task, while in SPAM, there is no need for task halts. Although SAGAT is claimed to be unintrusive (Endsley, 1995a), it is reasonable to think that stopping a simulation or task several times may result in some interference. Endsley, Sollenberger, Nakata, and Stein (2000) conducted a study to evaluate the interference caused by the application of

SAGAT and SPAM. Ten qualified controllers from five Air Route Traffic Control Centers (ARTCCs) in the U.S. subjectively rated the intrusiveness of the two methods. The mean ratings shows that they considered SAGAT to be more intrusive than SPAM to ATC simulations. At the same time, in real work environments, flight operation on the radar screen is never halted. Therefore, a simulation with SPAM administration would be more realistic compared to a simulation with SAGAT administration. In addition, it is easy, quick, and low-cost to use the SPAM technique. Because of these advantages, SPAM was chosen in the current study.

Task Performance

The criteria for good task performance usually vary from task to task and from people to people (Austin & Villanova, 1992). For example, in air-to-air combat simulations, the number of enemy fighter jets lost and the number of friendly forces killed were collected to indicate pilot performance (Endsley, 1995a). In dual-task driving scenarios, the probability of missing traffic signals and the mean reaction time to them were used to reflect drivers' performance (Strayer & Johnston, 2001).

In ATC scenarios, different types of data are collected to reflect task performance, such as cognitive performance, task scores, and peer-rating grades. Ackerman and Cianciolo (2002) measured memory perceptual speed, pattern perceptual speed, scanning perceptual speed, and complex perceptual speed to describe and compare subjects' task performance in TRACON simulations. Federal Aviation Administration (FAA) ATST is a dynamic low-fidelity simulation of the en-route ATC work environment (Nickels, Bobko, Blair, Sands, & Tartak, 1995). Variables such as error counts, arrival delay, and handoff delay are computed automatically to reflect subjects' performance (Federal

Aviation Administration, 1997). In the ASU ATC laboratory, experienced FAA air traffic controllers observe students' ATC simulations and then comprehensively evaluate their performance based on the subtask performance such as air traffic judgement, separation control, and coordination.

FAA Air Traffic Control Specialist Performance Measurement Database (ATC-SPMD) is a collection of techniques and data types that can validly and reliably describe air traffic controllers' performance in simulations (Hadley, Guttman, & Stringer, 1999). The purpose of ATC-SPMD is to combine effective ATC performance measurement techniques into a single source and to improve standardization of ATC performance parameters across different studies related to ATC, which enables comparisons of ATC performance data across these studies. According to four ATC environments (tower, TRACON, en-route, oceanic), ATC-SPMD is categorized into four subsets. The database consists not only of tools that measure controllers' performance, but also tools for measuring effectiveness and cognitive performance. With reference to controllers' performance, the tools focus on aspects such as traffic difficulty, operational errors, and radiotelephone communication. Flight safety and strategy efficiency are viewed as indicators of controllers' effectiveness. Cognitive performance can be reflected by variables such as planning, decision making, and strategy execution. ATC-SPMD allows a researcher to search and select performance parameters that are appropriate to different experimental conditions and serve a variety of purposes in different studies. In the current study, four types of simulation output data, such as handoff delay time, average time of aircraft in sector, operational errors, and commands issued, were chosen to indicate ATC students' simulation performance with reference to the ATC-SPMD. Handoff delay time measures the timeliness of establishing initial contact with an airplane and transfering it to the next sector. Average time of aircraft in sector indicates how efficiently an ATC student can expedite the traffic flow. Operational errors can reflect whether a student meets the flight safety goal. The data of commands issued describes the frequency of interactions between a student and the pseudo pilot, which provides information about the student's strategy execution, proficiency, and decision making.

Workload

To improve the design of human-machine systems, much attention has been paid to workload studies. According to Jex (1998), workload is the amount of physical and mental work that a person has to do in a task. Generally, an air traffic controller's goal is to guide airplanes to prevent accidents as well as expedite traffic flow in the air or on the ground (Nolan, 1994). To fulfill the goal, controllers need to do real-time decision making and communicate with pilots and other controllers, both of which are aspects of a controllers' mental workload and physical workload. As different controllers may have different ATC strategies and cognitive abilities, their physical and mental workload in the same scenario may also vary to some extent.

A series of workload measurements have been developed and studied so far. Subjective Workload Assessment Technique (SWAT) is an online workload self-rating technique (Reid & Nygren, 1988). SWAT includes three dimensions to assess time load, mental effort load, and stress load. During a simulation or task, subjects use three levels (low, medium, and high) to rate the three dimensions, which results in a total of 27 possible combinations of the ratings. Before assessing each subject's workload, the method requires subjects to respectively sort the 27 possible outcomes in a order that mirrors their ideas about increasing workload. Then each subject's workload data are respectively compared with this order to measure subjects' perception of the workload of a simulation or task. Instantaneous Self Assessment of workload (ISA) is similar to SWAT, and is also an online self-rating workload measurement (Brennan, 1992). Unlike SWAT, ISA measures how busy a subject is with one 5-point scale. In the scale, 1 means not busy and 5 means very busy. ISA shows up at fixed intervals during a simulation or task to measure workload, which helps detect the change of one subject's workload during the simulation or task. However, in a systematical comparison between SWAT and ISA, the two methods are considered to be disruptive in varying degrees (David & Pledger, 1995).

Some workload measurements were specially developed for ATC scenarios. For example, Air Traffic Workload Input Technique (ATWIT) is an online, subjective measurement that reflects the workload produced by the air traffic system (Stein, 1985). Specialized equipment called the Workload Assessment Keypad (WAK) is required to collect workload data during an ATC simulation. Subjects are required to rate their workload on a 7-point scale, in which 1 means low workload and 7 means high workload, at several time windows during the simulation. WAK records the rating and the rating input time. Although it can describe subjects' perception of the workload changes, the unidimensional scale in ATWIT is considered a limitation in obtaining more information about workload (Federal Aviation Administration, 2001).

NASA Task Load Index (NASA-TLX) form is an offline, self-rating technique with six subscales (Byers, Bittner & Hill, 1989). The six subscales respectively assess mental demand, physical demand, temporal demand, performance, effort, and frustration. Each subscale assessment has a 21-point scale, from 0 to 20. In mental demand, physical demand, temporal demand, effort, and frustration subscales, 0 means very low and 20 means very high. On the contrary, in performance subscale, 0 means perfect and 20 means failure. The total score of the six subscales is viewed as a workload indicator. NASA-TLX is widely accepted and applied in scientific research. It is a multi-dimensional measurement, which can also provide extra information about subjects' perception of workload in a simulation. According to Hart (2006), NASA-TLX is a tool considered to be reliable, sensitive and easy to apply. In the current study, the NASA-TLX was used to measure ATC students' workload during the high-fidelity ATC simulations.

Studies in Air Traffic Scenario Test

Durso et al. (2006) chose ATST to examine the relationship between SPAM and subjects' performance. In the ATST simulation, arrows represent airplanes. There is a bordered area on the simulation screen that represents the airspace where airplanes fly and maneuver. On the right side of the screen, information such as airspeed, heading, and flight level of an airplane are displayed. There are four gates at each side of the bordered area which are represented by "A", "B", "C", and "D". Two airports, which are represented by "e" and "f", are separately located in the bordered area. There are three speeds (fast, medium, and slow), three flight levels (gate departure level, mid level, and landing level), and eight directions for each airplane. Subjects are to efficiently and safely guide different airplanes from random starting locations to their destinations at the correct flight level, speed, and heading. For example, if the designated destination of an airplane is airport "e", the airplane should descend to the landing level, slow down to slow speed, and directly head to the appropriate runway. Subjects use a mouse to activate and control airplanes. Once an airplane shows up in the bordered area, it will not start to move until participants click on it with the mouse. In the study, students with no ATC experience were recruited to perform the en-route ATC-like simulations. Three variables were used to evaluate subjects' performance. The first is handsoff delay time: the latency between the moment an airplane shows up in the area and the moment a subject activates it with the mouse. The second variable is ATC errors, including hitting the border, guiding airplanes to the wrong destination and other errors. The third is average en-route delay time for airplanes. En-route time is how much time it takes for an airplane to move from its original location to its destination. Each airplane can spend the least time to get to its destination with the best heading and highest speed. Therefore, the delay is the difference between the two measurements of time.

In the study conducted by Durso et al. (2006), there were baseline conditions and SPAM conditions. Subjects performed the ATST simulations without any disruption in the baseline conditions, while they were required to answer SPAM questions during the simulations in the SPAM conditions. SPAM questions occurred every 2.5 to 3 minutes, and one question was asked each time. Researchers found no significant difference between the three performance variables (handsoff delay time, ATC errors, and average en-route delay time for airplanes) in the two conditions. The result suggests that the presence of SPAM may not affect subjects' performance in ATST simulations.

Pierce et al. (2008) found that in the previous study, although there is no statistical difference between performance in the two conditions, the data of all three performance variables shows that subjects tended to have better performance in the baseline conditions than in the SPAM conditions. Due to the lack of workload measurement, whether a "Ready" signal is enough to eliminate the impact of SPAM questions on task workload during the ATST simulation still remains unknown. In their study, Pierce et al. (2008) continued to use FAA's ATST as their experimental simulation. The subjects were all students without any ATC experience. Besides the handsoff delay time, en-route delay time, and ATC errors, variables such as the number of commands issued and the number of airplanes correctly handled were added to better describe task performance. Commands issued reflects how many times a subject changes the airplanes' status (heading, speed, and flight level) during a simulation. The number of airplanes correctly handled is the total number of airplanes that arrive at their destinations with the correct status in one simulation. Baseline conditions and SPAM conditions remained the same as in the Durso et al. study (2006). In addition, word shadowing conditions and list memory conditions were added to the experiment. Instead of being asked SPAM questions, subjects were required to repeat several words as soon as the words showed up after a "Ready" signal in word shadowing conditions. In list memory conditions, subjects would hear a list of words after the "Ready" signal. Then after an extra verbal signal, they were asked to repeat the words in the same order as they were presented. In word shadowing and list memory conditions, only cognitive resources are required for subjects to process the word information before repeating the words. So the purpose of adding these two conditions to the study is to learn whether SPAM questions have an impact on subjects' workload and performance in a way of simply consuming extra cognitive resources or in some more complex ways. Workload in the four conditions was measured according to the NASA-TLX form. In the study, significant differences were found between baseline and SPAM conditions in terms of handsoff delay time, the number of commands issued, and the number of airplanes correctly handled. This finding provides evidence that the application of SPAM may affect the subjects' primary task performance. Workload data was also analyzed but no significant difference was found, which suggests that the administration of SPAM questions did not add extra primary task workload to the subjects in the ATST scenarios.

Pierce (2012) then redesigned and improved the study. Instead of using both visual and verbal SPAM questions, only verbal SPAM questions were applied in this new study. As the information required for ATST tasks was perceived entirely visually by subjects through the simulation screen, visual SPAM questions were considered to be more intrusive than verbal ones. All ATST scenarios in the new study were redesigned to have the same difficulty and the same amount of workload. Pierce (2012) collected the number of commands issued from 21 seconds before a "Ready" signal to 21 seconds after the "Ready" signal, as it was hypothesized that the decrease in performance caused by the SPAM administration should be most obvious during the time immediately preceding and following the "Ready" signals. The time period then was equally divided into 6 smaller periods, each 7 seconds in duration. As there was no "Ready" signal in the baseline conditions, the number of commands issued over the equivalent time periods in the baseline conditions was collected. In Pierce's (2012) study, the four conditions (SPAM conditions, baseline conditions, word shadowing conditions, and list memory conditions) remained the same. Workload was measured using NASA-TLX forms.

The analysis revealed that there was a significant difference between the number of commands issued from 7 seconds after a "Ready" signal to 14 seconds after the "Ready" signal in the baseline and SPAM conditions. A significant difference was also observed between the number of commands issued from 14 seconds after a “Ready” signal to 21 seconds after the “Ready” signal in the two conditions. The results show that there is a decrease in subjects’ interactions with airplanes immediately following the SPAM questions in SPAM conditions. The analysis also revealed that the number of commands issued from 7 seconds after a “Ready” signal to 14 seconds after the “Ready” signal was significantly lower than the number of commands issued in all other time periods in the SPAM conditions. This provides evidence that the decrease in performance immediately following the “Ready” signal in SPAM conditions may be caused by the SPAM administration in the ATST tasks. In terms of the number of airplanes correctly handled, there was a significant difference between the two conditions, which suggests that subjects’ performance was affected by the administration of SPAM technique. However, no difference was found between workload measured by NASA-TLX form in the two conditions. This result shows that the administration of SPAM in the ATST simulation did not add extra task workload.

Although these studies have been conducted in the ATST simulation provide rich information about the relationship between the SPAM administration, workload, and task performance (Durso et al., 2006; Pierce et al., 2008; Pierce, 2012), no such study has been conducted using high-fidelity ATC simulations with subjects who have ATC experience. ATST is a simplified ATC simulation and there is a vast difference between ATST and high-fidelity simulations. For example, in ATST scenarios, subjects only need a mouse to control the airplanes, while subjects who perform tasks in high-fidelity ATC simulations have to issue commands to a pseudo pilot with standard radiotelephone communication language. In high-fidelity simulations, different types of airplanes have different speed ranges and ceilings. The difference of speed ranges between a Boeing 737 and a Boeing 747 is an example. However, in ATST, each airplane has only three speeds, three flight levels, and eight directions, which, to some extent, lowers the complexity and difficulty of an ATC simulation. The lack of airplane performance diversity in low-fidelity simulations can also cause a decrease in the SPAM question difficulty. For example, a SPAM question about speed in ATST simulations only has three potential answers (slow, medium, and fast). In general, the big gap between ATST and high-fidelity simulations is likely to lead to incorrect predictions about the relationship between SPAM, workload, and performance in high-fidelity ATC simulations with subjects who have ATC expertise and experience.

CHAPTER 3

METHOD

Participants

Sixteen Arizona State University students (*M* = 21.94 years old, *SD* = 0.77 years), 15 males and 1 female, took part in the study. All subjects major in Air Traffic Management and have ATC 433, En-route Operation & Procedure, course experience.

Study Design

Data was collected in two experimental simulation conditions in the ASU ATC simulation laboratory. All sixteen subjects performed the simulations in the two experimental conditions: baseline conditions and SPAM conditions. Between the two conditions, each one had a five-minute break. Two new simulation scenarios (Scenario A and Scenario B) were developed for the experiment (see Appendices C and D for complete scenario information). The order effect of the two conditions and the potential difference between the two simulation scenarios were fully counterbalanced by using each possible combination of the conditions and scenarios equally often (see Appendix E for the combination information).

In baseline conditions, each subject performed the ATC simulation once. One of the two scenarios was chosen. No SPAM questions were asked during the simulation in baseline condition. Upon completion of the simulation, subjects were required to file the NASA-TLX forms.

In SPAM conditions, each subject performed the ATC simulation once. The other one of the two scenarios was administered to make sure no subjects met the same scenario twice. Four SPAM questions were asked after “Ready” signals during a simulation (see Appendix F for the question list). The first SPAM question was asked three minutes after the simulation began. The remaining SPAM questions occurred about every three minutes. All of the SPAM questions were asked verbally by the researcher. After the simulation, subjects were asked to file the NASA-TLX forms.

Simulation

In the ASU ATC simulation laboratory, two new high-fidelity en-route ATC scenarios were built with medium workload and moderate difficulty in the Adacel system, with each simulation lasting 15 minutes. They were also designed to have the same difficulty and same amount of workload.

There were 14 airplanes in each of the scenarios. The destination of 10 airplanes was Phoenix Sky Harbor Airport (PHX) while the other four airplanes' destinations were not PHX. Subjects were asked to sequence the PHX inbound traffic. At the same time, they were to maintain positive control of all airplanes and ensure flight safety at all times (see Appendix D for scenario guide and aircraft information).

Performance Measurement

Based on the data that can be collected in the ATC laboratory, combined with the FAA ATC-SPMD (Hadley et al., 1999), four kinds of data were selected as the simulation performance variables.

- Handoff delay time: the total number of seconds that the airplanes crossed the sector boundary before being radar contacted and handed off.
- Average time of aircraft in sector: the average number of seconds an airplane stayed in the sector.
- Operational errors: the frequency of aircraft separation violations, radio

communication mistakes and other errors.

- Commands issued: the number of commands issued from one minute before a "Ready" signal to a certain point in time after it. This time period minus the time for SPAM question and answer in the time period is two minutes.

Workload Measurement

NASA-TLX form was also used in the experiment. Subjects were required to rate their workload after each simulation. Workload in the two simulation tasks was indicated by the scores of NASA-TLX forms.

CHAPTER 4

RESULTS

Software and Calculations

All of the data collected was analyzed in Statistical Product and Service Solutions (SPSS). An alpha level of .05 was used for all statistical tests in this study. In both baseline and SPAM conditions, the handsoff delay time of the vast majority of subjects was zero. As a result, the handsoff delay time was not analyzed and was removed from the study. Initial analyses showed that all of the other data groups have approximately normal distributions. Average time of aircraft in sector, operational errors, and NASA-TLX scores were analyzed using paired-sample *t*-tests. The number of commands issued was analyzed using repeated ANOVA, with conditions and time period as within-subject factors.

Workload

Table 1 includes subjects' overall NASA-TLX scores and scores of the six subscales. Marginal significant difference was observed between mental demand scores in baseline conditions ($M = 10.8$, $SD = 5.0$) and in SPAM conditions ($M = 12.3$, $SD = 3.7$), $t(15) = -2.12$, $p = .051$. The comparison of the overall NASA-TLX scores between baseline conditions ($M = 48.2$, $SD = 16.2$) and SPAM conditions ($M = 53.5$, $SD = 14.8$) revealed no significant difference, $p = .069$. No significant difference was found comparing all of the other subscales of NASA-TLX form between baseline conditions and SPAM conditions, all *p's* > .126.

Table 1

Mean and SD (in parentheses) of NASA-TLX scores in baseline and SPAM conditions

Condition	NASA-TLX	Mental Demand	Physical Demand	Temporal Demand	Performance	Effort	Frustration
Baseline	48.15	10.84	5.53	7.71	5.09	11.84	7.12
	(16.22)	(4.96)	(4.13)	(4.81)	(3.07)	(3.59)	(5.64)
SPAM	53.46	12.25	5.43	8.96	6.93	12.68	6.62
	(14.76)	(3.69)	(4.54)	(4.08)	(5.57)	(3.78)	(3.44)

Note. NASA-TLX = NASA Task Load Index form; SPAM = Situation Presence Assessment Method

Performance

The average time of aircraft in sector was compared between baseline conditions (*M*=771.3, *SD* = 8.36) and SPAM conditions (*M*=774.0, *SD* = 6.93) and no statistical significance was observed, $p = .132$. There was no significant difference between the number of operational errors in baseline conditions (*M*=2.69, *SD* = 1.74) and in SPAM conditions (*M*=2.75, *SD* = 1.77), $p = .903$.

Both the number of commands issued within one minute before the arrival of a "Ready" signal and the number of commands issued from the time point of the "Ready" signal to a certain time point after the signal in SPAM conditions were measured. The latter time period minus the time for answering a SPAM question that occurred during the time period was one minute. Since there was no "Ready" signal in the baseline conditions, the number of commands issued over the equivalent time periods in the baseline conditions were measured. As a result, there were 16 groups of data (see Appendix B for the group list). The data of the mean number of commands issued is depicted in Figure 1.

A repeated ANOVA with condition and time period as within-subject factors was conducted to compare the performance between the two conditions. There was a significant main effect of condition, $F(1,15) = 6.11$, $p = .026$, $\eta_p^2 = 0.29$. There was no significant main effect of time period, $p = .620$, and no significant condition × time interaction, $p = .151$.

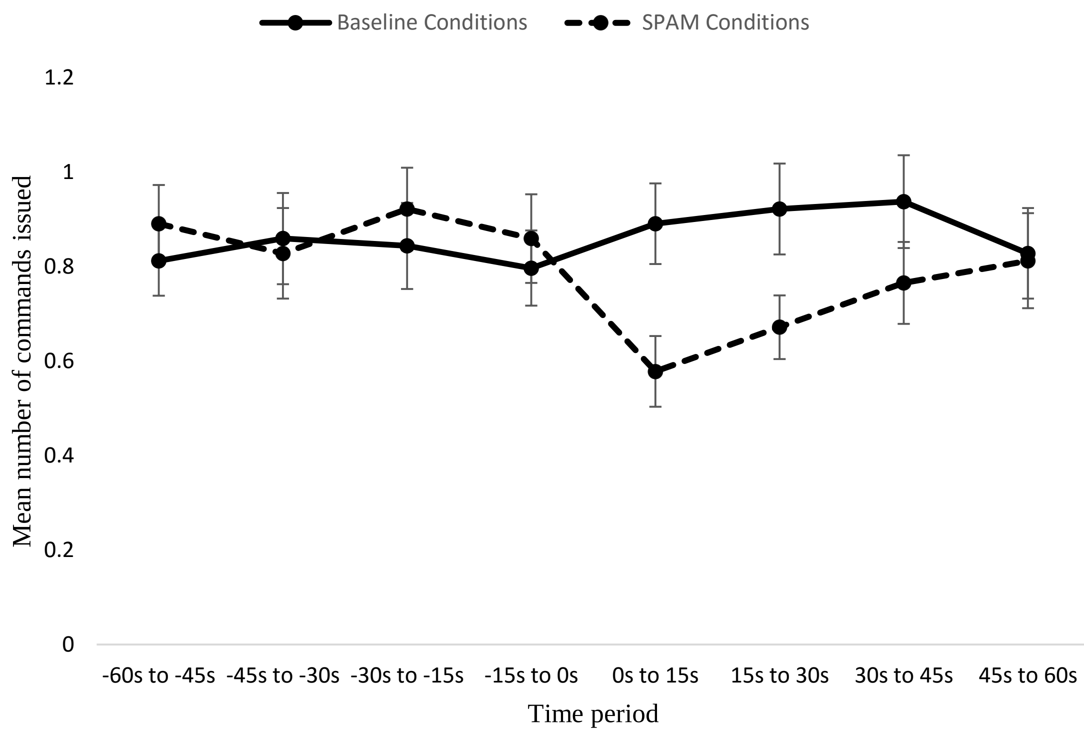

Figure 1. The mean number of commands issued before and after the "Ready" signals in baseline and SPAM conditions.

CHAPTER 5

DISCUSSION

In this study, the relationship between SPAM, task workload, and primary task performance was tested. To be specific, the study provides evidence to answer the question of whether the SPAM question administration during the high-fidelity en-route ATC simulation had an impact on ATC students' workload and simulation performance.

Simulation workload was indicated by NASA-TLX form scores in this study. There was no significant difference between the scores of overall workload in the baseline and SPAM conditions. No significant difference was observed between the scores of physical demand, temporal demand, performance, effort, and frustration in the two conditions. Meanwhile, marginal significant difference was observed between mental demand scores in the two conditions. The results show that although the SPAM administration may not affect the overall workload of the high-fidelity simulation task, there is a trend of higher mental workload in the SPAM conditions than in the baseline conditions. However, NASA-TLX form is offline and cannot reflect dynamic changes of workload during a simulation (Byers et al., 1989). Therefore, the change in workload immediately preceding and following the SPAM questions in this study remains unknown. In future studies, online workload measurements such as ATWIT could be applied simultaneously to collect more information about workload changes in SPAM conditions.

In terms of average time of aircraft in sector and operational errors, there was no significant difference between these performance variables in the two conditions. The results suggests that the simulation performance may not be affected by the administration of SPAM technique.

In terms of commands issued, the repeated ANOVA revealed a significant main effect of condition. The result shows that ATC students' performance in SPAM conditions was significantly poorer than that in baseline conditions, which agrees with the hypothesis, which is if SPAM is applied, ATC students' performance in high-fidelity en-route simulations would be poorer than that in baseline conditions. There was no significant main effect of time period, which means no significant difference between the number of commands issued in different time periods. The result of no significant condition and time interaction shows the performance data in the two conditions did not change over time in significantly different ways. However, a clear drop of the number of commands issued immediately after the "Ready" signal in SPAM conditions and a recovery trend of the commands issued after the signal in SPAM conditions can be observed.

To sum up, the study provides evidence that the administration of SPAM may, to some extent, affect ATC students' performance after the SPAM questions are asked in high-fidelity en-route simulations. The study also shows a trend of higher mental workload in the simulations with SPAM administration than without SPAM administration. The decrease in the number of commands issued after the "Ready" signal in SPAM conditions suggests that "Ready" signals may not be enough to eliminate dual-task interference caused by the SPAM questions and the decrease in primary task performance following the questions.

In the current study, the scenarios simulated en-route ATC operation with moderate difficulty and medium traffic flow, so it remains unknown whether the impact of the SPAM administration would be the same in en-route simulations with different levels of difficulty and workload. Four SPAM questions were asked in each simulation in SPAM conditions and the questions occurred every three minutes. Therefore, if the number of SPAM questions and the time interval between the questions change, whether the impact of SPAM administration on ATC students' workload and performance in en-route simulations will be the same remains unknown. Since en-route control is different from Tower and TRACON control, whether the impact of the SPAM administration will be the same in Tower and TRACON simulations is also unknown.

REFERENCES

Ackerman, P., & Cianciolo, A. (2002). Ability and task constraint determinants of complex task performance. *Journal of Experimental Psychology-Applied*, 8(3), 194-208.doi:10.1037//1076-898X.8.3.194

Austin, J. T., & Villanova, P. (1992). The criterion problem: 1917-1992. *Journal of Applied Psychology, 77*(6), 836-874. doi:10.1037/0021-9010.77.6.836

Australia Department of Defense, Defense Science and Technology Organization (2005). *An empirical study of the relationship between situation awareness and decision making* (Report No. DSTO-TR-1687). Retrieved from http://dspace.dsto.defence.gov.au/dspace/handle/1947/4318

Brennan, S. D. (1992). *An experimental report on rating scale descriptor sets for the instantaneous self-assessment (ISA) recorder* (CAD5 Report No. 92017). Portsmouth: DRA Maritime Command and Control Division.

Byers, J. C., Bittner, A. C., & Hill, S. G. (1989). Traditional and Raw Task Load Index (TLX) Correlations: Are Paired Comparisons Necessary? *Advances in Industrial Ergonomics and Safety*, 481-485.

Chiappe, D. L., Strybel, T. Z., & Vu, K. L. (2012). Mechanisms for the acquisition of situation awareness in situated agents. *Theoretical Issues in Ergonomics Science*, *13*(6), 625-647. doi:10.1080/1463922X.2011.611267

David, H., & Pledger, S. (1995). *Intrusiveness of On-Line Self-Assessment in ATC Simulation using Keyboard and Speech Recognition*. Retrieved from http://www.eurocontrol.int/eec/public/standard_page/DOC_Report_1995_004.html

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), 721-721. doi:10.1518/001872006779166316

Durso, F. T., & Dattel A. R. (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., Crutchfield, J., Nikolic, D. and Manning, C., (1998), *Situation awareness as a predictor of performance for en route air traffic controllers* (Report No. DOT/FAA/AM-99/3). Retrieved from http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/1990s/media/am99-03.pdf

Durso, F. T., Truitt, T. R., Hackworth, C. A., Crutchfield, J. M., Ohrt, D. D., Nikolic, D., Moertl, P. M., & Manning, C. A. (1995). Expertise and Chess: a pilot study comparing situation awareness methodologies. In D. J. Garland & M. R. Endsley (Eds.), *Experimental Analysis and Measurement of Situation Awareness* (pp. 295-303). Daytona Beach, FL: Embry-Riddle Aeronautical Press.

Endsley, M. R. (1990). Predictive Utility of an Objective Measure of Situation Awareness. *Human Factors and Ergonomics Society Annual Meeting Proceeding, 34*(1), 41-45. doi: 10.1177/154193129003400110l0

Endsley, M. R. (1995a). Measurement of situation awareness in dynamic systems. *Human Factors, 37*(1), 65-84. doi:10.1518/001872095779049499

Endsley, M. R. (1995b). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*(1), 32-64. doi:10.1518/001872095779049543

Endsley, M. R. & Rodgers, M. D. (1998). Attention distribution and situation awareness in air traffic control. *Air Traffic Control Quarterly, 6*, 21-44.

Endsley, M. R., Sollenberger, R. L., Nakata, A., & Stein, E. S. (2000). *Situation awareness in air traffic control: enhanced displays for advanced operations* (Report No. DOT/FAA/CT-TN00/01). Retrieved from http://hf.tc.faa.gov/technotes/dot-faa-ct-tn00-01.pdf

Federal Aviation Administration. (1997). *Effects of video game experience on computer-based air traffic controller specialist, air traffic scenario test scores: Office of aviation medicine reports* (Report No. DOT/FAA/AM-97/4). Retrieved from http://www.faa.gov/data_research/research/med_humanfacs/oamtechre ports/1990s/media/am97-04.pdf

Federal Aviation Administration. (2001). *Investigating the validity of performance and objective workload evaluation research (POWER)* (Report No. DOT/FAA/AM -01/10). Retrieved from http://www.cami.jccbi.gov/aam-400A/Abstracts/ 2001 /FULL%20TEXT/0110.pdf

Hadley, G. A., Guttman, J. A., & Stringer, P. G. (1999). *Air traffic control specialist performance measurement database* (Report No. DOT/FAA/CT-TN99/17). Retrieved from http://www.tc.faa.gov/its/worldpac/techrpt/cttn99-17.pdf

Hart, S. G. (2006). Nasa-task load index (NASA-TLX); 20 years later. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, *50*(9), 904-904.

Holmgvist, K. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford; New York: Oxford University Press.

Jeannot, E., Kelly, C., & Thompson, D. (2003). *The development of situation awareness measures in ATM systems* (Report No. HRS/HSP-005-REP-01). Retrieved from https://www.eurocontrol.int/sites/default/files/content/documents/nm/safety/safety-the-development-of-situation-awareness-measures-in-atm-systems-2003.pdf

Jex, S. M. (1998). *Stress and job performance: Theory, research, and implications for managerial practice*. Thousand Oaks, Calif: Sage Publications.

Marois, R., & Todd, J. J. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, *428*(6984), 751-754. doi:10.1038/nature02466

Nickels, B. J., Bobko, P., Blair, M. D., Sands, W. A., & Tartak, E. L. (1995). *Separation and control assessment (SACHA) final job analysis report*. Bethesda, MD: University Research Corporation.

Nolan, M. S. (1994). *Fundamentals of air traffic control*. Belmont, CA: Wadsworth.

Paige, B. L., & 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. doi:10.1016/j.ssci.2012.06.019

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, *116*(2), 220-244. doi:10.1037/0033-2909.116.2.220

Pierce, R. S. (2012). The effect of SPAM administration during a dynamic simulation. *Human Factors: The Journal of Human Factors and Ergonomics Society*, *54*(5), 838-848. doi:10.1177/0018720812439206

Pierce, R. S., Vu, K. L., Nguyen, J., & Strybel, T. Z. (2008). The relationship between SPAM,workload, and task performance on a simulated ATC task. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, *52*(1), 34-34.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128-2148. doi:10.1016/j.clinph.2007.04.019

Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. *Advances in Psychology*, *52*(C), 185-218. doi:10.1016/S0166-4115(08)62387-0

Stanton, N. A., Salmon, P. M., Rafferty, L. A., Walker, G. H., Baber, C., & Jenkins, D. P. (2013). *Human factors methods: A practical guide for engineering and design, 2nd edition.* doi:10.1080/00140139.2014.948659

Stein, E.S. (1985). *Air traffic controller workload: An examination of workload probe.* (Report No. DOT/FAA/CT-TN84/24). Atlantic City, NJ: Federal Aviation Administration Technical Center. Retrieved from http://hf.tc.faa.gov/ technotes /dot-faa-ct-tn84-24.pdf

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science*, *12*(6), 462-466. doi:10.1111/1467-9280.00386

Strybel, T. Z., Vu, K. L., Kraft, J., & Minakata, K. (2008). Assessing the situation awareness of pilots engaged in self spacing. *Proceedings of the 52nd Annual Meeting of the Human Factors & Ergonomics Society*, Human Factors Society, CA.

Taylor, R. M. (1989). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. *Proceedings of the AGARD AMP Symposium on Situational Awareness in Aerospace Operations*, North Atlantic Treaty Organization Advisory Group for Aerospace Research and Development, Seuilly-sur Seine, France

Taylor, R. M., & Selcon, S. J. (1994). Situation in mind: Theory, application, and measurement of situational awareness. In R. D. Gilson, D. J. Garland, & J. M. Koonce (Eds.), *Situational awareness in complex settings* (pp. 69-78). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Weick, K. (1990). The vulnerable system: An analysis of the Tenerife air disaster. *Journal of Management*, *16*(3), 571-593.

Woods, D. D., & Sarter, N. B. (1991). Situation awareness - A critical but ill-defined phenomenon. *The International Journal of Aviation Psychology*, *1*(1), 45-57. doi:10.1207/s15327108ijap0101_4

APPENDIX A

IRB APPROVAL FORM

APPROVAL: EXPEDITED REVIEW

Mary Niemczyk
Polytechnic School - Aviaton Programs
480/727-1595
Mary.Niemczyk@asu.edu

Dear Mary Niemczyk:

On 3/2/2016 the ASU IRB reviewed the following protocol:

Type of Review:	Initial Study
Title:	The effect of the situation presence assessment method (SPAM) on workload and task performance during high-fidelity ATC simulation
Investigator:	Mary Niemczyk
IRB ID:	STUDY00004034
Category of review:	(7)(b) Social science methods, (7)(a) Behavioral research
Funding:	None
Grant Title:	None
Grant ID:	None
Documents Reviewed:	• Revised on 0302 Recruitment-script.pdf, Category: Recruitment Materials; • Revised on 0302 HRP-503a-_PROTOCOL_SocialBehavioralV Chao Zhang.docx, Category: IRB Protocol; • NASA Task Load Index form.pdf, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • Revised on 0302 HRP-502a-Social behavior Consent form for Chao Zhang's study.pdf, Category: Consent Form; • SPAM questions list.pdf, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • CITI Program IRB social&behavior research training Report.pdf, Category: Other (to reflect anything not captured above);

The IRB approved the protocol from 3/2/2016 to 3/1/2017 inclusive. Three weeks before 3/1/2017 you are to submit a completed Continuing Review application and required attachments to request continuing approval or closure.

If continuing review approval is not granted before the expiration date of 3/1/2017 approval of this protocol expires on that date. When consent is appropriate, you must use final, watermarked versions available under the "Documents" tab in ERA-IRB.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Sincerely,

IRB Administrator

cc: CHAO ZHANG
Michael Pearson
Robert Nullmeyer
CHAO ZHANG

APPENDIX B

COMMANDS ISSUED DATA GROUP LIST

Commands Issued Data Group List

1	The number of commands issued from 60 s before the “Ready” signal in SPAM conditions to 45 s before the “Ready” signal in SPAM conditions
2	The number of commands issued from 45 s before the “Ready” signal in SPAM conditions to 30 s before the “Ready” signal in SPAM conditions
3	The number of commands issued from 30 s before the “Ready” signal in SPAM conditions to 15 s before the “Ready” signal in SPAM conditions
4	The number of commands issued from 15 s before the “Ready” signal in SPAM conditions to 0 s before the “Ready” signal in SPAM conditions
5	The number of commands issued from 0 s after the “Ready” signal in SPAM conditions to 15 s after the “Ready” signal in SPAM conditions
6	The number of commands issued from 15 s after the “Ready” signal in SPAM conditions to 30 s after the “Ready” signal in SPAM conditions
7	The number of commands issued from 30 s after the “Ready” signal in SPAM conditions to 45 s after the “Ready” signal in SPAM conditions
8	The number of commands issued from 45 s after the “Ready” signal in SPAM conditions to 60 s after the “Ready” signal in SPAM conditions
9	The number of commands issued in baseline conditions from 60 s before the equivalent time point of the “Ready” signal to 45 s before the equivalent time point of the “Ready” signal
10	The number of commands issued in baseline conditions from 45 s before the equivalent time point of the “Ready” signal to 30 s before the equivalent time point of the “Ready” signal
11	The number of commands issued in baseline conditions from 30 s before the equivalent time point of the “Ready” signal to 15 s before the equivalent time point of the “Ready” signal
12	The number of commands issued in baseline conditions from 15 before the equivalent time point of the “Ready” signal s to 0 s before the equivalent time point of the “Ready” signal
13	The number of commands issued in baseline conditions from 0 s after the equivalent time point of the “Ready” signal to 15 s after the equivalent time point of the “Ready” signal
14	The number of commands issued in baseline conditions from 15 s after the equivalent time point of the “Ready” signal to 30 s after the equivalent time point of the “Ready” signal
15	The number of commands issued in baseline conditions from 30 s after the equivalent time point of the “Ready” signal to 45 s after the equivalent time point of the “Ready” signal
16	The number of commands issued in baseline conditions from 45 s after the equivalent time point of the “Ready” signal to 60 s after the equivalent time point of the “Ready” signal

APPENDIX C

SCENARIO MAP

Scenario A

Scenario B

APPENDIX D

SCENARIO GUIDE AND AIRCRAFT LIST

Scenario Guide and Aircraft List

1. You will be given two scenarios on Albuquerque Center (ZAB). Both scenarios are with the same workload. Similar procedures to that used in ATC433 are used.
2. Aircraft will be coming in on several streams or arrival flows from the eastern portion of the sector. Aircraft already in the sector, within the lateral and vertical boundaries, are assumed to be on your frequency. You do not need to establish initial contact, or use your callsign with the first transmission to the aircraft, if the aircraft was within the boundary when the scenario started.
3. Pay close attention to the arrival stream and the direction of the aircraft to identify its route.
4. Flight strips are provided for your convenience.
5. The scenario lasts 15 minutes. Not all aircraft will exit the airspace in the 15 minutes.
6. You must maintain positive control and separation at all times. You are to use the appropriate phraseology and procedures used in the 7110.65.
7. Aircraft must be 8 NM in trail, constant or increasing, by the time the aircraft reaches SLIDR, if the aircraft is landing at PHX.
8. PHX arrivals must also cross SLIDR at and maintain FL290.
9. You may issue step down descents. You may not issue a descent below FL300 until the aircraft is abeam or past GUP.
10. TUS arrivals on the EAGUL, or passing over SLIDR, shall remain at the flight level they entered the airspace.
11. Speed changes can only be done in increments no greater than ± .03, one time, per aircraft.

Aircraft list:

Scenario A

Callsign	Ac Type	Dest
AAL611	A320	JFK
AAL965	B737	PHX
ASA912	B738	PHX
ASH1278	CRJ2	PHX
DAL260	MD83	PHX
LIFTER31	C17	PHX
N192ME	C750	DEN
N5301H	B722	LGB
SWA443	B733	TUS
UAL112	A333	PHX
UAL396	A320	PHX
UAL6011	B752	PHX
USC401	LJ35	PHX
WJA1004	B738	PHX

Scenario B

Callsign	Ac Type	Dest
AAL216	A333	PHX
AAL510	B737	BOS
AAL554	A320	PHX
AAY190	MD83	PHX
ASH113	CRJ2	PHX
LIFTER55	C17	PHX
N232AB	C750	SLC
N252JR	B722	LAX
SWA100	B738	PHX
SWA437	B733	TUS
UAL1050	B752	PHX
UAL272	A320	PHX
USC515	LJ35	PHX
WJA810	B728	PHX

APPENDIX E

THE COMBINATIONS OF CONDITIONS AND SCENARIOS

All Possible Combinations of the Conditions and Scenarios

Combination	First simulation	Break	Second simulation
1	Scenario A in baseline condition	5-min break	Scenario B in SPAM condition
2	Scenario B in baseline condition	5-min break	Scenario A in SPAM condition
3	Scenario A in SPAM condition	5-min break	Scenario B in baseline condition
4	Scenario B in SPAM condition	5-min break	Scenario A in baseline condition

APPENDIX F

SPAM QUESTION LIST

Spam Question List

For Scenario A:

1. What is the relative location of UAL112 towards N192ME right now?
2. Can AAL611 climb right now?
3. Which one has the higher altitude, USC401 or AAL965?
4. Will UAL6011 and ASA912 be in conflict if no further action is taken?

*For all of the en-route simulations that are run with Scenario A in SPAM conditions, the four questions above are asked in random order in each of the simulations.

For Scenario B:

1. What is the relative location of UAL272 towards WJA810 right now?
2. Can N232AB climb right now?
3. Which one has the higher altitude, SWA437 or AAL554?
4. Will UAL1050 and AAL190 be in conflict if no further action is taken?

*For all of the en-route simulations that are run with Scenario B in SPAM conditions, the four questions above are asked in random order in each of the simulations.