Spring 2011

Analysis of different phases of a commercial flight using radio call response times, workload, situation awareness and fatigue ratings

Ahmed Faruk Diken
University of Iowa

Recommended Citation

Diken, Ahmed Faruk. "Analysis of different phases of a commercial flight using radio call response times, workload, situation awareness and fatigue ratings." MS (Master of Science) thesis, University of Iowa, 2011.
https://doi.org/10.17077/etd.l4cnxqrp

Follow this and additional works at: https://ir.uiowa.edu/etd
Part of the Industrial Engineering Commons

ANALYSIS OF DIFFERENT PHASES OF A COMMERCIAL FLIGHT USING RADIO CALL RESPONSE TIMES, WORKLOAD, SITUATION AWARENESS AND FATIGUE RATINGS

by

Ahmed Faruk Diken

A thesis submitted in partial fulfillment
of the requirements for the Master of
Science degree in Industrial Engineering
in the Graduate College of
The University of Iowa

May 2011

Thesis Supervisor: Associate Professor Thomas Schnell

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

MASTER'S THESIS

This is to certify that the Master's thesis of

Ahmed Faruk Diken

has been approved by the Examining Committee
for the thesis requirement for the Master of Science
degree in Industrial Engineering at the May 2011 graduation.

Thesis Committee:
Thomas Schnell, Thesis Supervisor

Geb Thomas

Yong Chen

To My Parents

ACKNOWLEDGMENTS

The completion of this thesis would not have been possible if it weren't for the tireless efforts and assistance of a number of people at the Operator Performance Laboratory at the University of Iowa. First of all, I am deeply grateful to my thesis supervisor, Associate Prof. Thomas Schnell, for giving me the opportunity to be his research assistant and for guiding me throughout my graduate studies and thesis project. I would also like to thank Mathew Cover, one of the lead research engineers at the Operator Performance Laboratory, who assisted in troubleshooting issues with hardware that was required during the experiments, for giving feedback during the development phase of the experiments, and also for being the assistant experimenter and spending long hours of the night crammed in a simulator rather than relaxing at home. Thirdly, I would like to thank Nicholas Anderson, the simulation expert at the Operator Performance Laboratory, for turning the difficult task of designing the simulation into an easy problem to solve. The hours he spent with me in editing the Air Traffic calls, which were crucial for the experiments, and the insight he provided while designing the study, are truly appreciated. I would also like to thank Michael Yocius, the software engineer at the Operator Performance Laboratory, who helped in solving software related issues and provided the application that extracted response times from the raw data that was collected during the experiments. I would also like to acknowledge Darin Hoover and Martin Schlegel who helped in this study as the assistant experimenter. I thank them both for their time in simulating the Air Traffic Control during the experiments. I would also like to acknowledge Carlton Richey for the time he spent in helping to prepare the

Boeing 737 simulator for the experiments and resolving any technical issues that came up during the course of this study. I thank all the Operator Performance Laboratory staff for creating and maintaining one of the best research laboratories of the University of Iowa.

Finally, I thank my parents who always motivate me with their exemplary parenthood to work hard and keep moving forward.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Statement of the Problem

Pilots are subject to varying levels of stress, workload, and fatigue during long flights. During different phases of a commercial flight, pilots are engaged in multiple tasks which include going through checklists, checking conditions at their destination, communicating with Air Traffic Control and dealing with other flight related tasks. The amount of work varies from the earlier stages until the end of the flight. It is not well understood how changes in the amount of workload can affect a pilot's ability to engage with important tasks that relate to safety of flight. The work shown in this thesis focused on the level of engagement displayed by flight crew as a function of level of workload. The principal hypothesis was that very low levels of workload may lead to crew disengagement and sub-optimal levels of performance. The degree to which pilots remain alert and are fatigued during a commercial flight is also not established in a concrete way.

Proposed Solution

One of the primary objectives during any flight, whether commercial or combat related, is assuring the safety of the aircraft and everyone on board. Many variables affect the safety of a given flight. The most obvious of these variables can be listed as the condition of the aircraft, the training and experience of the pilots flying the aircraft and external conditions such as weather. In addition to these, the mental and physiological state of the pilots can have a direct impact on the safe operation of an aircraft. Many aircraft accidents have been attributed to human factors related issues such as fatigue. Other conditions that can be included in the human factors list are boredom and low levels of workload. Different phases of flight can exert varying levels of workload on pilots. For example, the workload that a pilot endures during take-off or landing, and the workload experienced during auto-pilot assisted flight (such as in level flight) are at vastly different levels. Furthermore, during level flight, it is likely that the pilots will experience boredom because of a lack of cognitively stimulating tasks. Fatigue can also set in as the flight progresses, and the severity of this can vary depending on how many assignments the pilot has already fulfilled in a given day. Flights at night are especially prone to elicit extreme levels of fatigue and disengagement in flight crews. All of this can affect the pilot's situation awareness and his decision making ability during an unexpected and potentially dangerous event. Being able to measure and assess a pilot's workload and how this may affect decision making can lead to the development of technology that can assist the pilots during flight by employing means that ensure pilot's awareness of the state of the aircraft as well as other conditions such as traffic and weather.

The latent stress and workload that the pilot experiences during a flight cannot be detected with the current set of avionics on board an aircraft. However, objective and subjective measurements from carefully designed experiments can shed light on how workload, fatigue and boredom play a factor in the decision making capabilities of the pilot. Determining a set of effective measures and a way to measure them without interfering with the pilot's mode of operation of the aircraft can lead to incorporating the outcomes of such research into avionics. Such results can lead to the aircraft and the pilot becoming an integrated joint system and a higher level of flight safety can be achieved. Conducting this type of research with the aid of high fidelity simulators and flight scenarios resembling real circumstances a pilot endures can lead to interesting results. The objective of this thesis is to cover some ground on this issue by conducting a simulation exercise involving actual pilots and realistic flight scenario based on a night-time transcontinental flight.

The University of Iowa, located in Iowa City, Iowa, is home to the Operator Performance Laboratory (OPL), which is a division of the Center for Computer Aided Design research facility. According to the OPL’s website, the facility houses a Boeing 737 simulator, a Joint Strike Fighter (JSF) simulator, a general aviation simulator, two Aero Vodochody L-29’s, and an A36 Bonanza aircraft. This study used the Boeing 737 simulator. Specifically, the study statistically analyzed the changes in response times to radio calls during different phases of flight such as take-off, level flight and landing as well as the reported workload, situation awareness and fatigue by pilots using post-run surveys.

CHAPTER 2

BACKGROUND

This chapter is a review of the literature on factors that can affect pilot performance. These factors can be classified into two broad categories: factors related to the mental and cognitive state of the pilot and factors related to the aircraft. Pilot related factors may include situation awareness, workload, and fatigue, pilot's level of training, issues related to personal life that may have an adverse affect on performance, crew dynamics, job satisfaction, the physical and environmental factors of the aircraft, and the airworthiness of the aircraft in terms of the mechanical, electronic and other systems on board. In this chapter, the focus of the literature review will be on the mental and cognitive factors. More specifically, these will include situation awareness, workload, and fatigue.

Situation Awareness

One of the factors that can affect pilot performance is called situation awareness. According to Mica R. Endsley (2010), situation awareness (SA) is defined as "the internalized mental model of the current state of the flight environment" (p. 12-1). More specifically, it is described as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" (as cited in Endsley, Selcon, Hardiman, & Croft, 1998, p. 1). Endsley model is described as one of the most popular SA models in the field of ergonomics (Salmon et al., 2009, p.490). It is made up of three levels, each level representing a stronger awareness and comprehension of the condition of the aircraft than the one before it. The levels are perception of the elements in the environment, comprehension of the current situation, and projection of future status. In the first level of SA, the pilot perceives the important internal elements related to the aircraft as well as the external environmental elements such as traffic, weather and terrain. In the second level of SA, the pilot is not only described as being aware of the different elements affecting the aircraft, but also understanding their significance and how they impact the objectives of the flight. For instance, at level two, the pilot needs to understand what a fuel warning light means for his flight and how this affects the airworthiness of the aircraft. Experienced pilots are usually at the third level of SA, in which they not only understand the current state of the aircraft but also how the different elements may project themselves into the future of the flight (Endsley, 2000, p. 7; Endsley, 2010, p. 12-3).

In addition to being categorized into levels, SA is also categorized based on type. Types of SA include geographic SA, spatial SA, system SA, environmental SA, and in the case of military aircraft, tactical SA. Geographic SA is related to the location of the aircraft, other aircraft that may be in the vicinity, features of the terrain, airports in the area, waypoints, and other elements related to landing, take-off and taxi procedures (Endsley, 2010, p. 12-3). Spatial/temporal SA is related to the pilot’s comprehension of the aircraft’s altitude, attitude, heading, velocity, flight path related elements, projected landing time and other similar variables. System SA refers to the pilot’s awareness of the radio and autopilot settings, state of Air Traffic Control (ATC) communications, fuel related issues, malfunctions related to the operation of the aircraft and how these may affect the airworthiness of the aircraft, as well as any other variables related to the performance of the aircraft. Environmental SA is related to weather related issues that may affect the safe operation of the aircraft at the present time as well as in the future and the pilot’s awareness of instrument flight rules in case visibility does not allow the pilot to use visual flight rules (Endsley, 2010, p. 12-4).

Figure 1 below shows the decision making model, discussed by Mica R. Endsley in his paper entitled “Theoretical Underpinnings of Situation Awareness: A Critical Review”. The three levels that were discussed earlier are depicted in the figure. According to Endsley (2000), SA is an important component in the decision making process, but it is not the sole determinant of the decision’s integrity. Other factors, such as training, stress, workload and system capabilities have an impact on the integrity of the decision being made. As noted by Endsley, however, this does not take away from the importance of SA as a critical factor that affects decision making (p. 8).

Factors Affecting Situation Awareness

There are multiple factors that affect SA. One of these factors is the limited attention capacity of a pilot. Because of the large number of tasks and amount of information a pilot needs to remain aware of during the entire course of a flight, actively paying attention to each instrument and warning can increase the stress and workload level. This can also reduce SA depending on the strategy that the pilot employs to divide attention between high and low priority tasks. The National Transportation Safety Board (NTSB) indicated that, according to accident reports, about 31% of accidents were related to poor SA due to attention problems (Endsley, 2010, p.12-4).

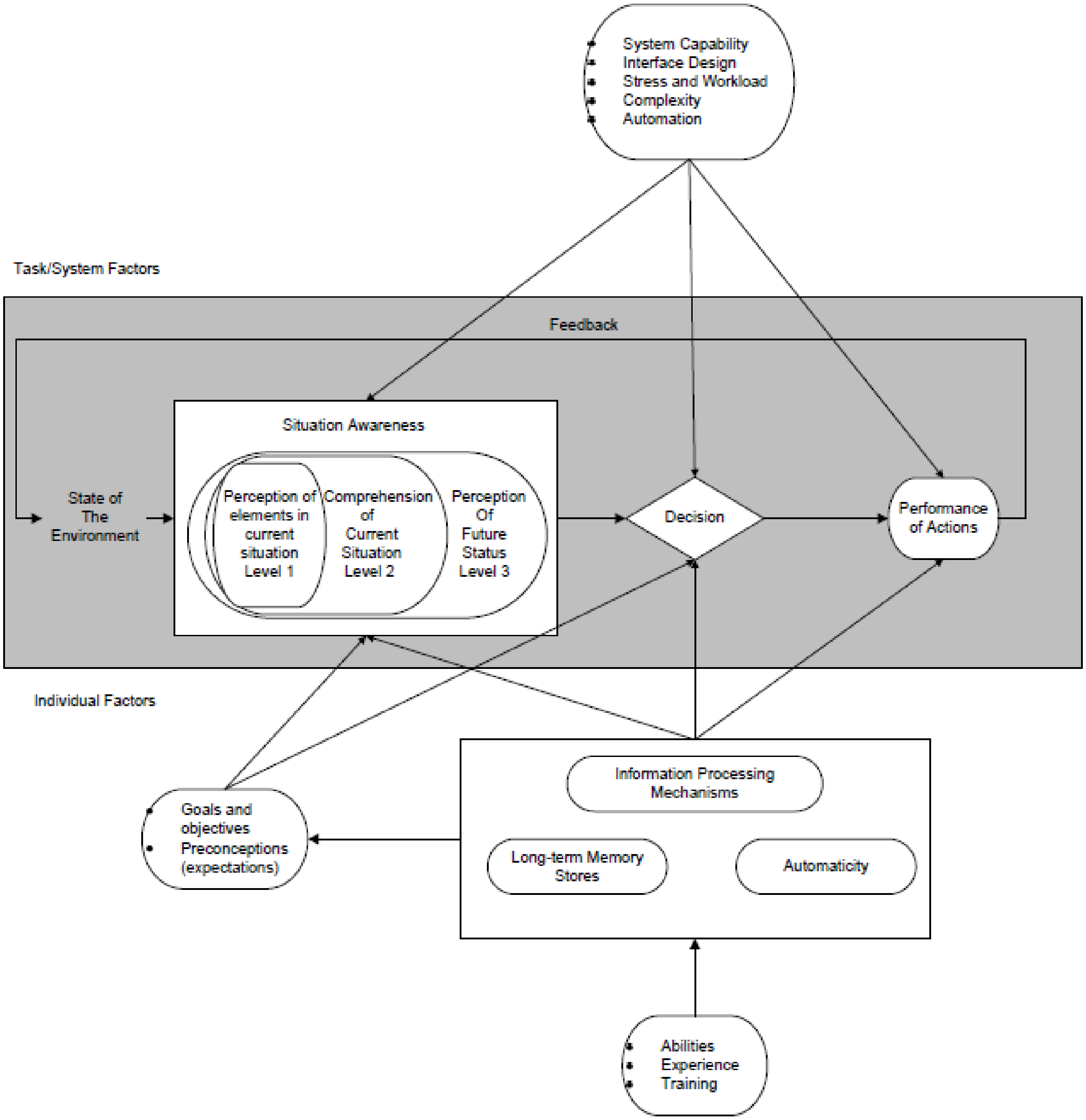


Figure 1. Endsley's Model of Situation Awareness

Source: Mica R. Endsley, Daniel J. Garland, 2000, *Situation Awareness Analysis and Measurement*, Mahwah, NJ, Lawrence Erlbaum Associates, Inc., Publishers; John A. Wise, V. David Hopkin, Daniel J. Garland, 2010, *Handbook of Aviation Human Factors*, Boca Raton, FL, CRC Press

As SA is a construct that requires the pilot to populate and maintain a mental model of a situation, it follows that working memory is an important factor that can impact the level of SA. The second and third levels of SA in which the pilot is required to comprehend the meaning of the information being gathered from the different systems and project that information into the future of the flight can be extremely taxing on working memory (Endsley, 2000, p. 12). In dividing attention in complex situations, Endsley (2000) noted that according to research in this field, “attention to information is prioritized based on how important that information is perceived to be” (p. 13). Working memory limitations can lead to more severe setbacks for novice decision makers than for experienced decision makers, who may have developed strategies to deal with this problem (p. 14).

Long term memory and mental models help in dealing with the complexity of information that decision makers have to deal with, especially in the case of experienced pilots. Endsley (2000) described that, “operators develop internal models of the systems they operate and the environment in which they operate” (p. 16). Unlike novice aircrews who may have to rely solely on working memory to perceive, comprehend and project new information into the future state, experienced pilots can utilize mental models in long term memory as well as working memory to make quicker decisions (Endsley, 2010, p. 12-6). Wrong mental models can also be negative for SA. Errors may occur when the situation is interpreted incorrectly, or if the mental model is erroneous (Endsley, 2000, p. 17). Automaticity, or automatic processing, in which an often repeated task is completed without attention, is another variable that can impact SA. It is achieved when oft repeated and habitual acts are carried out with very low level of attention. Although automatic processing can help relieve the load on attention, subtle differences can be missed by the decision maker, leading to errors which can sometimes be detrimental (Endsley, 2010, p. 12-7).

Thus far, the factors that are directly related to the decision maker's mental and cognitive limitations such as working memory, long term memory, mental models and attention were discussed. Stress is another important factor that can affect SA. Monica Martinussen and David R. Hunter (2010) mentioned that, according to some theories, stress is "the result of factors or elements that have a negative impact on the individual" (p. 127). Different types of stressors can influence a pilot's SA and how attention is divided across tasks. For instance, a person's character traits and how they respond to stressful situations, issues related to family life and the work environment can all have an impact on the ability of the decision maker to achieve a high level of SA (Endsley, 2010, p. 12-8). Endsley (2010) categorized stressors into two types. The first type is physical stressors such as noise, vibration, lighting, atmospheric conditions, fatigue, boredom and the temperature of the environment. The second type is social and psychological stressors which include fear or anxiety, importance or consequences of events, uncertainty, mental load, and time pressure (p. 12-8). Endsley (2010) further explained that these stressors can affect SA in different ways. Some amount of stress may actually help the decision maker pay attention to important elements, but too much of it can lead to reduction or loss of SA. More specifically, high stress situations can lead pilots to divert attention from central tasks to more peripheral tasks and lose control of the aircraft due to a loss in SA (p. 12-8). Another factor that affects SA is workload, which will be discussed in more detail later.

Situation Awareness Measurement Techniques

Salmon et al. (2009) explained that there are a number of different measurement techniques because the definition of situation awareness is not universal (p. 491). They have classified the measurement methods into several categories. The first, freeze probe techniques, require the experimenters to freeze a simulation after a certain task and administer a questionnaire to calculate an SA score for that particular task or scenario. The score is calculated by considering the feedback of the participant as well as the state of the system. One of the methods under this category is SAGAT, which uses the three level model of situation awareness developed by Endsley (p. 491). It is an objective measure, which involves asking the participant questions about the circumstances surrounding the task at the time the simulation was frozen. The questioner then scores SA based on the correctness of the answers. The nature of SAGAT's administration is also its main disadvantage. Due to the random nature of the pauses, participants do not have time to prepare themselves for the questions. This may have an adverse effect on the answers provided, thereby leading to an erroneous SA score (Endsley, Selcon, Hardiman, & Croft, 1998, p. 82).

Real-time probe techniques include scoring methods that are administered while the simulation is running. Subject matter experts score the participant's situation awareness based on the content of responses and reaction times. The advantage of this technique over the freeze-probe technique is its non-intrusive nature in the sense that the continuity of the simulation is not disrupted. A popular scoring method under this category is the situation present assessment method, also referred to as SPAM (p. 491).

Self-rating techniques, unlike the two discussed so far, are subjective assessments that are typically administered after the simulation has ended. The participants use a scale to rate their perceived level of situation awareness for the tasks they carried out during the simulation (p. 491). Situation awareness rating technique, also referred to as SART, is one of the most popular and thoroughly tested subjective SA rating scales (Endsley, Selcon, Hardiman, & Croft, 1998, p. 83; Jones, 2000, p.118). There are ten generic SA constructs that SART measures (Jones, 2000, p. 118). These are described in Table 1.

Table 1. SART Construct Definitions

SA Construct	**Definition**
Instability of situation	Likeliness of situation to change suddenly
Variability of situation	Number of variables which require one's attention
Complexity of situation	Degree of complication (number of closely connected parts)
Arousal	Degree to which one is ready for activity (sensory excitability)
Spare Mental Capacity	Amount of mental ability available to apply to new variables
Concentration	Degree to which one's thoughts are brought to bear on the situation
Division of Attention	Amount of division of attention in the situation
Information Quantity	Amount of knowledge received and understood
Information Quality	Degree of goodness or value of knowledge communicated
Familiarity	Degree of acquaintance with situation experience

Source: Mica R. Endsley, Daniel J. Garland, 2000, *Situation Awareness Analysis and Measurement,* New Jersey, Lawrence Erlbaum Associates, Inc.

These ten generic constructs can be clustered into three broad categories which are attentional demand, attentional supply and understanding. Attentional demand includes instability of situation, variability of situation, and complexity of situation. Attentional supply includes arousal, spare mental capacity, concentration, and division of attention. Finally, understanding encompasses information quantity, information quality and familiarity (p. 118). The constructs are scored on a seven point scale where one is low and seven is high.

The individual scores can be used to calculate a single score for SA. The formula for this is given below:

$$\text{SA (calc)} = \text{Understanding} - (\text{Demand-Supply})$$

Equation 1. Situation Awareness Formula (p. 119)

Observer-rating techniques are conducted during simulations or training sessions to assess the situation awareness of the participant by the subject matter expert. The difference between this method and the real-time probe technique is that in the former, the observer is the one who scores the situation awareness without feedback from the participant, while in the latter; it is the participant who answers questions while the simulation is being run. Although this characteristic makes observer-rating techniques the most non-intrusive, it is also questioned whether an observer can make the best assessment of a subject's situation awareness (Salmon et al., 2009, p. 492).

The performance measures technique uses aspects of task performance as a means to score situation awareness. Finally, process indices make use of data gathered from processes that are believed to help participants build situation awareness while carrying out a task. An example of this is eye tracking data. This kind of data, when analyzed in conjunction with participant's behavior during different tasks, can be used to assign a score for situation awareness (p. 492).

Pilot Workload

Most researchers who studied pilot workload agree that there is not a single, agreed upon definition of pilot workload (Hart, Staveland, 1988, p. 139; Rehmann, Stein, Rosenberg, 1983, p. 297; Watson, Ntuen, Park, 1996, p. 487). Sandra G. Hart (1988) defined workload as “a hypothetical construct that represents the cost incurred by a human operator to achieve a particular level of performance” (p. 140). The interaction of a number of factors describes the level of workload that is experienced by a pilot. These factors include the pilot’s skill level, the task that is being performed and the circumstances surrounding the task being performed (p. 140). Other definitions of pilot workload have been developed as well. Lloyd Hitchcock, Samira Bourgeois-Bougrine and Phillippe Cabon (2010) described pilot workload in their paper entitled “Pilot Performance”. According to the authors, pilot workload is defined as “the ratio of available resources and the required resources during the task. This means that a given task will not produce the same workload level for different operators (depending on their experience of this task), or even for the same operator (depending on his state during the task)” (p. 14-4). In this definition, the term “resources” refers to those required “for the attention used for the perception, the reasonable decision-making, and action” (p. 14-4). In another research article by Barry H. Kantowitz and John L. Campbell (1996) entitled “Pilot Workload and Flightdeck Automation”, pilot workload is described as follows:

> Pilot workload is defined (Kantowitz, 1988a) as an intervening variable, similar to attention, that modulates or indexes the tuning between the demands of the environment and the capacity of the operator. As an intervening variable, workload cannot be directly evaluated or observed. Instead, it is a conceptual, multifaceted construct that must be inferred from changes in observable data” (p. 118).

Francis T. Durso and Amy L. Alexander (2010) described mental workload as "the relation between the function relating the mental resources demanded by a task and those resources supplied by the human operator" (p. 219). The attention pilot workload has received in research implies that it is a critical measure that has wide implications. The use of workload measurements have, in part, provided guidance in designing cockpits such that two instead of three pilots can now fly large commercial aircraft. It has also contributed to improved technological advances in Air Traffic Control (p. 229).

Kantowitz and Campbell (1996) have built a model that describes how different factors can affect pilot workload (p. 119). Figure 2 illustrates this model. The model shows that there are certain factors that can increase workload and there are factors that can decrease it. Factors such as pilot skill, feedback from the automated system and system reliability decrease pilot workload, while task demands, environmental demands and pilot fatigue increase pilot workload. The model also shows that factors interact with each other and the sum of their impact affects workload (p. 120). It has been shown that many aviation accidents that occurred during complex task performance were related to operator workload (p. 119). It is also highlighted that too much or too little workload can lead to errors of commission or omission (Kantowitz & Campbell, 1996, p. 120; Durso & Alexander, n.d., p. 219). Kantowitz and Campbell (1996) also drew special attention to the importance of low workload, which may be ignored when designing automated systems to reduce high workload. They discussed the conclusion of Kantowitz and Casper in a study conducted in 1988 in which the authors remarked on automation thus: "It would be ironic if our current efforts to measure pilot workload succeeded, only to be faced with a new generation of aircraft where pilot workload was so low that nobody bothered to measure it at all" (as cited in Kantowitz & Campbell, 1996, p. 120). This is one of the reasons why this study will not only try to understand how pilots respond to high workload situations, but also to low workload situations such as level flight where autopilot is usually engaged.

The methods used to measure workload are as varied as the definitions of workload that have been described. Valerie J. Gawron (2008) listed 36 different methods to measure workload. Among those that are most reliable, some of them employ the use of decision trees to assign a score for workload (p. 151). The Cooper-Harper Rating Scale is one of the notable workload measuring scales that uses a decision tree. It is, however, geared towards rating aircraft handling characteristics (p. 162). The Bedford workload scale is a variation of the Cooper-Harper rating scale and is more tailored towards assessing the workload of tasks. It was developed by trial and error with the help of test pilots at the Royal Aircraft Establishment at Bedford, England (p. 160).

The scale uses ten workload levels. These are described in the decision tree diagram in Figure 3. Sandra G. Hart and Lowell E. Staveland (1988) developed a more rigorous workload scale, called the NASA Task Load Index, referred to as NASA-TLX. Each task is scored using a scale with six components (p. 146). These components are listed as mental demand, physical demand, temporal demand, performance, effort, and frustration level (p. 170). A set of pairwise comparisons are used to determine which of the constructs were more important for the level of workload experienced while executing a task. An overall score is then calculated for each task using the weighted scores (p. 171). Typically, the workload scales are used in an intrusive manner, where the experimenter asks the subject what his perceived level of workload is during the performance of a task.

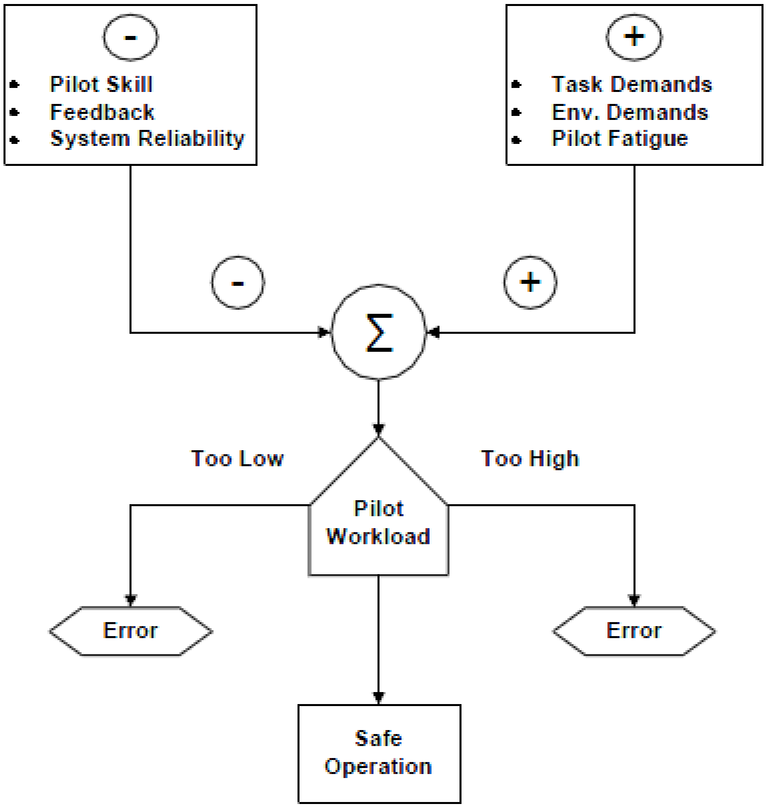


Figure 2. Model of Pilot Workload

Source: Raja Parasuraman, Mustapha Mouloua, 1996, *Automation and Human Performance*, New Jersey, Lawrence Erlbaum Associates, Inc.

During this study, however, the subjective assessment was performed using a survey which was administered after the experiment was finished. This approach was preferred because the experiment was designed to reflect the actual conditions of a flight, which requires the cockpit to remain sterile from conversations that are unrelated to the flight.

Pilot Fatigue

In general, fatigue is used to refer to a state of diminished capacity to work. It is borne out of prolonged activity, but is also influenced by psychological, socioeconomic and environmental factors (Brown, 1994, p. 298). Brown (1994) noted that fatigue can occur when a person is not able to fulfill the performance goals set for or by him/her but must continue working because of job requirements, or because not doing so may lead to deleterious consequences (p. 299). This describes the nature of cruise flight in modern aircraft, especially on nighttime transcontinental and transoceanic trips. On such trips, flight crews are exposed to long durations of work in an environment that can easily induce fatigue with no option to stop their task until the aircraft is parked at the destination. Others have defined fatigue from a physiological perspective. R. F. Soames Job and James Dalziel (2001) described fatigue thus:

> Fatigue refers to the state of an organism's muscles, viscera, or central nervous system, in which prior physical activity and/or mental processing, in the absence of sufficient rest, results in insufficient cellular capacity or systemwide energy to maintain the original level of activity and/or processing by using normal resources (p. 469).

Desmond and Hancock (2001) categorized fatigue into two kinds: active fatigue and passive fatigue. Active fatigue occurs during long tasks that require continuous activity while passive fatigue occurs during tasks that require monitoring but no direct activity at all times. In this sense, passive fatigue is related to vigilance. Active fatigue is experienced more commonly during ground transportation with motorists. Passive fatigue can be experienced by commercial pilots since they are not required to hand-fly their aircraft at all times, and rely on autopilot during a considerable part of flight (p. 455).

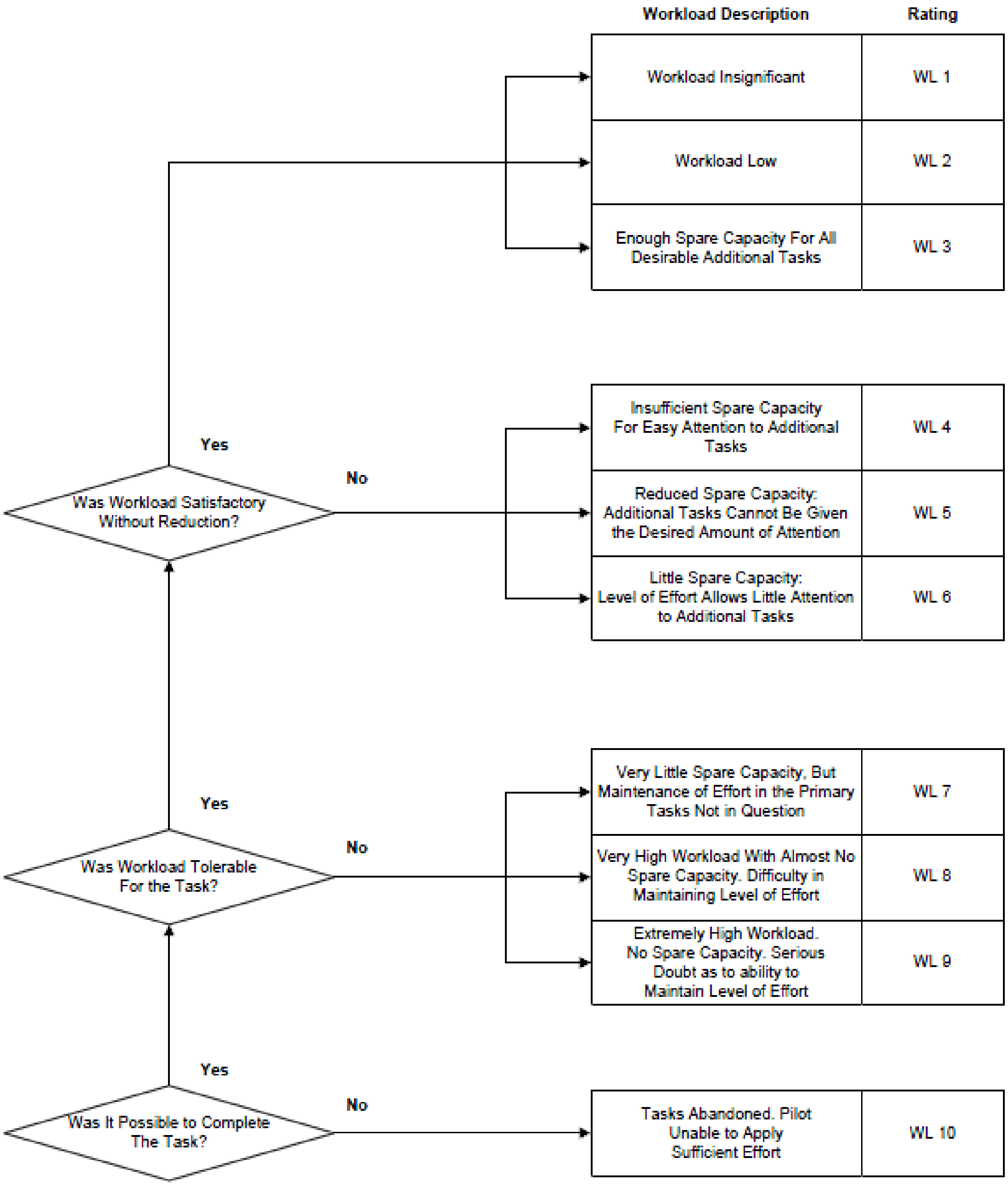


Figure 3. Bedford Workload Scale Decision Tree

Source: Valerie J. Gawron, 2008, *Human Performance, Workload, and Situational Awareness Measurement Handbook*, Florida, CRC Press

As described by the pilot workload model of Kantowitz and Campbell (1996), pilot fatigue is a significant contributor to increased pilot workload. Aviation accidents due to pilot fatigue have been a persistent and catastrophic problem that has not been adequately addressed. John A. Campbell (2004) pointed to the fact that reports on accident statistics and reports from pilots identify fatigue as a significant problem in aviation operations (p. 86). Desmond and Hancock (2001) also cited accident reports from the National Highway Traffic Safety Administration that over 1,500 fatalities took place due to drowsiness caused by fatigue between 1989 and 1993. In regards to aviation, Lyman and Orlady found that 3.8% of aviation incidents were related to fatigue between 1976 and 1980 (as cited in Desmond & Hancock, 2001, p. 456). The National Transportation Safety Board (n.d.) reported on its website that there has been a concern about fatigued personnel who are engaged in critical tasks in the aviation industry in particular and the transportation industry in general. This issue has been listed on NTSB’s Most Wanted Transportation Safety Improvements list since 1990. Although aircraft accidents are not as common as accidents in other modes of transportation, they cause significant losses both in terms of human lives as well as the cost of the equipment that is ruined. Therefore, any step that can be taken to minimize accidents is welcomed by all stakeholders.

According to a CNN (2009) report, 250 fatalities in air accidents were attributed to pilot fatigue over the past 16 years. Several fatal crashes in recent years have been attributed to pilot fatigue. One of the most notable accidents of recent history is the Colgan Air flight, operating to Buffalo, NY from Newark, NJ on February 12th, 2009. According to the National Transportation Safety Board (2010) report on the accident, the flight crashed with 49 people onboard. All were killed including one person on the ground that was inside the residential unit the aircraft crashed into (p. 20). The NTSB (2010) also reported that the pilot who was in command of the aircraft commuted from Seattle, Washington to Newark, NJ the night before the accident. The investigators concluded that the crew did not get enough rest the night before the flight, which was identified as one of the factors that may have adversely affected the management of the aircraft when the stall occurred. Other factors such as lack of training, and not paying attention to the task at hand were also cited as reasons for the crash (p. 119). As reported on the Christian Science Monitor (2009), one of the causes of pilot fatigue is simply due to the fact that pilots are working longer hours than ever before which includes commuting to their place of work. The current Federal Aviation Agency regulations, which were first implemented when flying was not a common choice as it is now, do not take current trends into consideration. The current adverse economic conditions in general, and in the airline industry in particular, add more challenges on scheduling pilots for flights while making sure airlines do not run over their budgets.

Fatigue Due to Sleep Loss

One of the common causes of pilot fatigue is sleep loss, especially during long flights that coincide with late night hours. Many studies have been conducted in order to understand sleep loss and how it affects cognitive abilities. Research conducted at the Loughborough University’s Sleep Research Laboratory by J. Harrison and J. A. Horne (1999) reveal that sleep loss can lead to a loss in innovative thinking, flexible decision making, and can impair temporal memory (p. 128). Harrison and Horne (1999) point out that there is mounting evidence to suggest that one night without sleep can have a negative impact on a person’s ability to plan and think flexibly. The same researchers conducted a study in which the consequences of 36 hours of sleep loss were observed on ten healthy individuals. The researchers used a marketing decision making game and a critical reasoning test in order to measure how decision making capability was affected after the participants were deprived of sleep. They compared their performance with a group who played the same game without being deprived of sleep. The study was conducted for two days, and participants played the game and took the critical reasoning tests on both days. The purpose of the critical reasoning test was to ensure that any difference in participant’s performance in the marketing test was not due to information acquisition, but was because of impaired decision making ability (p. 128). At the conclusion of this study, the researchers found that scores in the critical reasoning test across the two groups did not have a significant difference. However, on the second day, after 30 to 36 hours of sleep deprivation, significant differences in performance between the sleep deprived and non sleep deprived groups during the marketing game were observed. The study showed that sleep deprived participants were unable to respond correctly to the changes in the game which was reflected in an increase in the number of production errors and a decrease in profitability (p. 136). Harrison and Horne (1999) explained that the participants started having difficulty in flexible and innovative thinking, which translated into the participants not being able to respond to rapidly changing events, leading ultimately to the collapse of the game (p.141). Based on their findings, the researchers recommend that “people working for extended periods of time, who are required to make decisions necessitating flexibility, and the ability to update plans in the light of new information, and presented under rapidly changing situations should avoid sleep loss beyond 32-36 hours" (p. 142). The type of working environment that Harrison and Horne describe fits that of an airplane cockpit, in which a pilot may experience situations that warrant quick decision making and planning ability. Goldman, McDonough, and Rosemond (1972) found that "junior doctors were more hesitant and showed less focused planning during a surgical operation" (as cited in Harrison and Horne, 2000, p. 236). A similar study by Nelson, Dell' Angela, Jellish, Brown, and Skaredoff (1995) revealed that "anesthesia residents who had no more than 30 minutes of sleep during a night on call had impaired innovative thinking and verbal fluency, whereas complex convergent tasks remained intact" (as cited in Harrison and Horne, 2000, p. 236).

Research has also been conducted to understand how shift changes, age, individual differences, and altitude in the case of pilots, can affect performance in combination with sleep loss. It is described that sleep and wakefulness is regulated by two neurobiological processes, namely the homeostatic process and circadian rhythm. The former is described as the process to keep the person awake during the day and asleep during the night. It is driven by what is called the "biological clock" of the brain, which keeps track of the time of day. The latter process is responsible for balancing between the time one spends sleeping with the time when one is awake (Van Dongen, 2006, p. 1140; Costa, 2010, p. 11-2). These two processes are synchronized under normal circumstances, when one spends the night asleep and works during the day. However, working night shifts can disrupt this balance. Although people who are on a permanent night shift schedule can adapt to this change over time, those who work night shifts more sporadically will not achieve a balanced state between the two processes. This translates into a person experiencing sleepiness during work hours which could have an adverse affect on productivity as well as safety (p. 1142). To recreate the conditions that a pilot experiences in terms of sleep loss would be difficult in an experiment. Such an experiment would require the pilots to continue on the same on-the-job sleep schedule during their days off when they are expected to rest and recover before their next work cycle. The best alternative where some fatigue can still be induced may be to schedule the experiments during evening hours so that pilots would run a night flight simulation during the time when they might normally be resting.

Fatigue Rating Scale

The Crew Status Survey, which was originally designed with 20 statements describing fatigue states, is a subjective survey designed to assess the workload and fatigue levels of air crews. The Crew Status Survey contains two sections, one for fatigue and the other for workload (Gawron, 2008, p. 173).

Table 2. Description of Fatigue Levels in the Crew Status Survey

	Subject Fatigue
1	Fully Alert; Wide Awake; Extremely Peppy
2	Very Lively; Responsive, But Not at Peak
3	Okay; Somewhat Fresh
4	A Little Tired; Less Than Fresh
5	Moderately Tired; Let Down
6	Extremely Tired; Very Difficult to Concentrate
7	Completely Exhausted; Unable to Function Effectively; Ready to Drop

Source: Valerie J. Gawron, 2008, *Human Performance, Workload, and Situational Awareness Measurement Handbook*, Florida, CRC Press

For this study, the section for fatigue was utilized without the section for scoring the level of workload. The Bedford workload scale described earlier was used to score the level of workload. Table 2 lists the seven levels of fatigue defined in the Crew Status Survey.

The Use of Reaction Times in Human Factors Studies

According to Gawron's (2008) research, reaction time is defined as "the time elapsed between stimulus onset and response onset" (p. 30). Reaction times have been used in a wide variety of research, including studies that involve driving simulations. In a study to understand the effects of age and mental workload in driving, Hiroshi Makishita and Katsuya Matsunaga (2008) used reaction times to a buzzer under five conditions which included drivers sitting in a stationary car, drivers engaged in mental calculations while sitting in a stationary car, during driving with no mental calculations, during driving with mental calculations and while driving in a public road (p. 568).

The study showed the reaction times for older participants during driving while engaged in mental calculations were longer when compared to younger drivers (p. 572). In another study, which somewhat parallels the objectives of the study at hand, Ping-Huand Ting, Jiun-Ren Hwang, Ji-Liang Doong and Ming-Chang Jeng (2008) looked at the effects of fatigue on highway driving using reaction times to assess sustained attention. The participants were required to respond to a visual stimulus in the display every 2 km and their reaction time was measured during the driving task (p. 449). Highway driving can be a monotonous task because of a lack of stimuli which can lead to impaired performance due to decreased vigilance (p. 448). In this sense, flying on an automated flight deck poses similar challenges, especially during level flight when not much is happening. The results of the study revealed that extended driving duration increased variability in reaction times which reflected an increase in fatigue as the driving task progressed (p. 451). A similar driving simulator study was conducted by Vincent Cantin, Martin Lavalliere, Martin Simoneau and Norman Teasdale (2009) who studied the effects of age and driving complexity. This study compared the reaction times of young and old participants who were asked to respond to an auditory task when driving in different scenarios (p. 764). The complexity of the driving task was varied during the study using three scenarios. The study found that mental workload increased with increasing complexity of the task and older drivers were found to have a greater increase in mental workload as complexity of the driving tasks increased than younger drivers did (p. 768).

CHAPTER 3

EXPERIMENTAL DESIGN AND HYPOTHESES

Methodology

In order to study the differences between phases of flight in terms of response times, workload, situation awareness and fatigue, the OPL conducted a simulation in a fixed base Boeing 737 flight simulator involving 15 pilots. The study included a flight from Seattle Tacoma International Airport (KSEA) to Chicago O’Hare International Airport (KORD). The details of the flight were extracted from an actual American Airlines flight which took place on May 10th, 2010. Details were provided on FlightAware. The flight path is represented by the blue line in Figure 4.

.

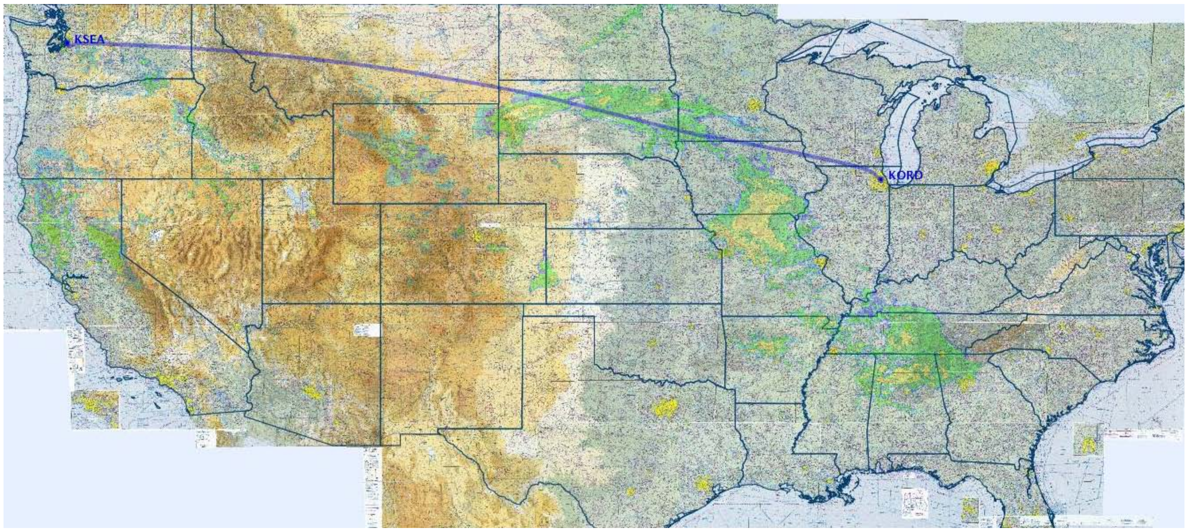


Figure 4. Flight Path for the Proposed Simulation (flightaware.com)

Source: Taken from http://flightaware.com/live/flight/AAL1238

In order to simulate the effect of fatigue that pilots may experience during long work days, all pilots were scheduled to participate in the experiment in the late evening hours and they were instructed to avoid drinking caffeinated drinks such as coffee during the day so that their system will be clear of any stimulant that may interfere with the integrity of the experiment. In effect, the simulation represented a full mission night flight from taxi and takeoff to parking at the destination gate.

After arriving at the OPL, subjects watched an orientation video and signed consent forms. The video contained a description of the experiment as well as instructions on how to program the Flight Management System (FMS) and use the autopilot to change the altitude and speed of the aircraft. The video was used to familiarize the subjects with the set up of the experiment, the experimental tools that were going to be used during the experiment, as well as the responsibilities that the experimenters had during the simulation. In addition, the pilots were asked to fill out a brief survey about their flying experience (number of years, type of aircraft flown, etc.). After the orientation, the pilots were given a tour of the simulator and final questions were answered. The pilots were also provided with airport diagrams for Seattle Tacoma and O’Hare International Airports. The experiment was conducted without any pauses in order to resemble an actual flight.

The lead experimenter ensured the proper functioning of all computers and sensors during the simulation and was in charge of tagging the data and noting down the time and nature of any events of interest that could influence the data analysis. The assistant experimenter controlled Air Traffic Control communications.

At the end of the simulation, the pilots were asked to take a survey to rate their workload, situation awareness and fatigue during different phases of flight. The experimental design team made a conscious decision not to administer any questionnaires during the flight as this may have stimulated the respondent thereby affecting the level of engagement, which was the observational variable of interest. The data gathered from these surveys were analyzed in conjunction with the data collected on response times to ATC calls. A copy of the survey is available in Appendix A.

Experimental Hypotheses

In this section, the expected hypotheses are discussed. As mentioned earlier, the purpose of this experiment is to determine how fatigue affects and task engagement in terms of reaction times, workload and situation awareness. In order to observe these differences, several hypotheses were tested as shown in Table 3.

Table 3. Experimental Hypotheses

Difference Between:	**Response Times to ATC Calls**		**Workload**		**Situation Awareness**		**Fatigue**	
	Null	**Alt.**	**Null**	**Alt.**	**Null**	**Alt.**	**Null**	**Alt.**
Level Flight - Takeoff/Ascent	0	≠0	0	≠0	0	≠0	0	≠0
Descent/Landing - Takeoff/Ascent	0	≠0	0	≠0	0	≠0	0	≠0
Descent/Landing - Level Flight	0	≠0	0	≠0	0	≠0	0	≠0

The data collected from taxi, takeoff and ascent to 37000 feet (when level flight begins) is treated as one homogeneous block of workload with a relatively wakeful pilot.

This also allows the collection of a sufficient number of ATC call responses. The four following stages were labeled as level flight in Seattle Center, Salt Lake Center, Minneapolis Center, and Chicago Center. During these stages, there was no change in the altitude of the aircraft. Also, the frequency of calls from ATC is less dense when compared to airspace surrounding large metropolitan airports. Typically, the calls from ATC are restricted to changing radio frequencies; a very simple task that involves read back of ATC instructions, adjusting the radio dials to the assigned frequency and letting the new Air Traffic Controller know that the aircraft is in the airspace of the center. The other stages of data collection include descent to 9000 feet, and final descent to the runway for landing. During these final stages of flight, it is expected that the pilot would become re-engaged and workload would be higher than in level flight as preparations for landing take place. The statistical analysis following the data collection looked at the differences in these stages using reaction times as well as reported workload, situation awareness and fatigue levels.

Participants

The pilots who volunteered for this study spanned a wide range of ages and experience levels. The youngest participant was 21 and the oldest participant was 64. The median age of participants was 40 and the average age was 42.1. The standard deviation of age among subjects was 14.7. Because the pool of subjects in the vicinity of the University of Iowa, where the experiments were conducted, was not as large as would be desired and many were not airline pilots, the requirements were not extreme. At minimum, the pilots were required to have a commercial pilot’s license with instrument rating, which means that the pilots could land the aircraft under conditions with limited visibility using the indicators on board the aircraft. Such circumstances would include foggy conditions.

Out of the 15 pilots who participated, all had experience flying single engine aircraft, four had experience flying multi engine aircraft, four had experience flying jet engine aircraft, and three had experience flying turboprop aircraft. Eleven of them carried commercial licenses, while four had private licenses. Thirteen had flown in a simulator before while it was the first time for two pilots. Figure 5 displays the years of experience each participant had with different kinds of aircraft.

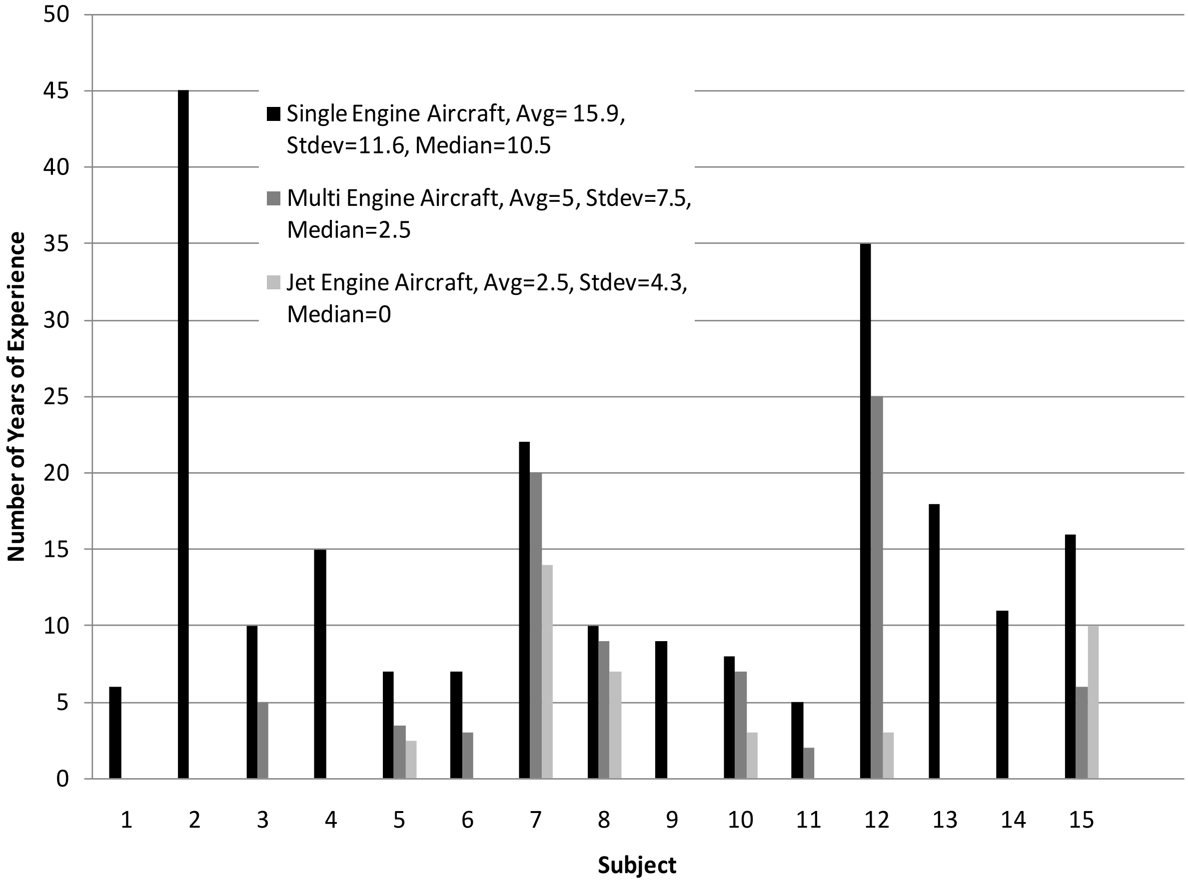


Figure 5. Flying Experience of Participants

Experimental Tools

In this section, the experimental tools that were used in the simulation are described in detail. These include the Boeing 737 simulator, the audio recording devices and the data collection applications.

Boeing 737 Simulator

The Boeing 737-800 simulator at the OPL is a fixed-base flight simulator with glass cockpit displays, five outside visual projectors, functioning mode control panel (MCP) with autopilot and auto-throttle, and standard Boeing 737 controls. A total of 11 computers in a complex network structure, and a host of software applications make up the Boeing 737 simulator. Among the computers used, one is dedicated to operate the Primary Flight Display (PFD), which shows the aircraft’s attitude, speed and altitude. Another computer is used to operate the Navigation Display (ND), which displays the flight plan and the relative position of the aircraft on its flight path along with waypoints and airports. Another computer is used to run the Flight Management Computer and display the Control Display Unit (CDU), which is used to program the flight plan. Another computer is used to run the simulation software—Microsoft Flight Simulator 2004—and display the Engine Instruments Cluster (EICAS), the overhead panel and the Master/Caution Panel on three separate displays. Five computers are used to run the software that projects the simulation onto a dome. Finally, a computer is used to take the inputs from controls operated by the pilot, such as the yoke, flaps, rudders, and throttles, and outputs data that changes the state of the aircraft on all the relevant computers. The overhead panel of the simulator has been equipped with a touch screen that allows the pilot to interact with all of the features present in a Boeing 737 overhead panel. The touch screen allows the flexibility of displaying the overhead panel of any airplane. Figure 6 shows a picture of the simulator at the OPL. Figure 7 shows the simulator with the complete setup of the experimenters' stations and a participant during a study.

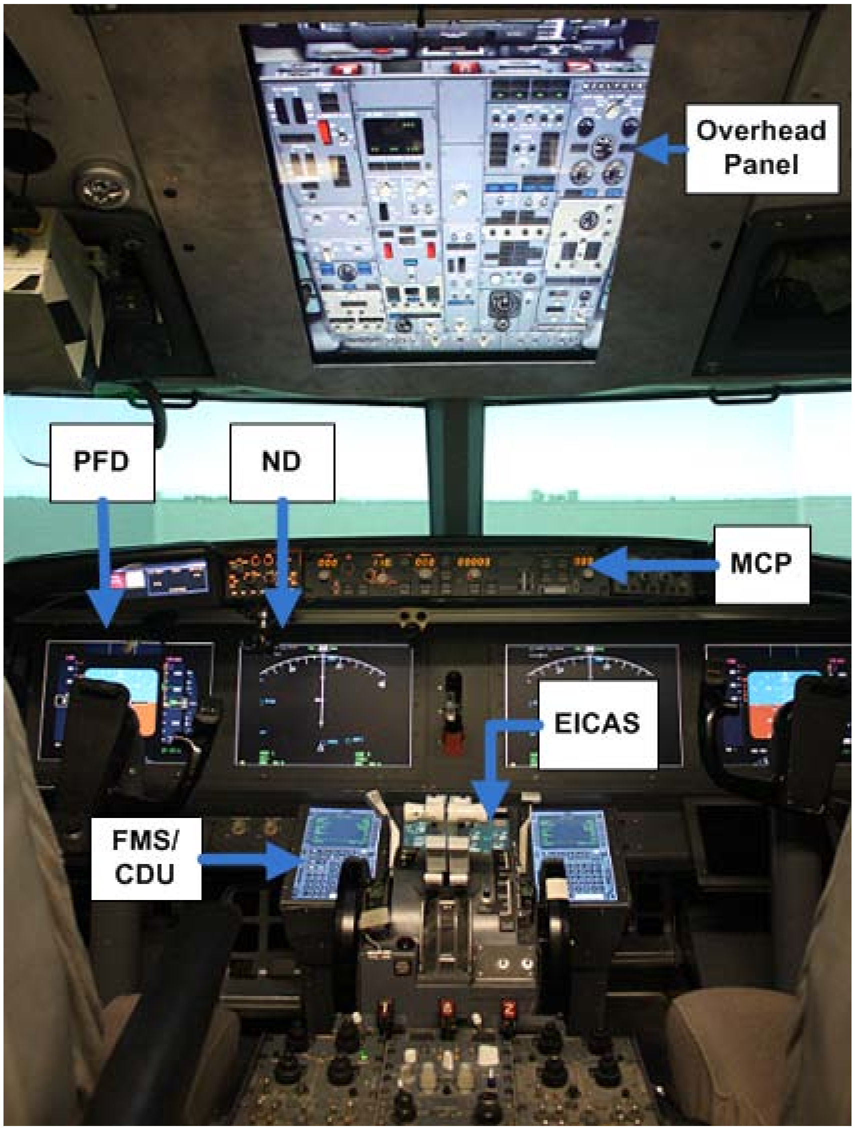


Figure 6. Boeing 737-800 Simulator at the OPL

Figure 7. Boeing 737 Simulator with the Stations of the Experimenters

Microsoft Flight Simulator 2004

The flight simulation software that was used for the experiment is Microsoft Flight Simulator 2004. The software comes with 24 aircraft, which includes a Boeing 737-700. The software allows the player to select from pre-programmed weather options including normal and severe weather conditions and allows the experimenter to set up timed or random failures on the engines as well as other equipment on the aircraft. A typical simulation has Air Traffic Control (ATC) calls as well. The ATC calls for the experiment were pre-recorded using the simulation software and edited using Adobe Soundbooth. The audio files were played by the experimenter at pre-determined times.

Subjective Survey Data Collection

The subjective rating scales that have been described so far can be disruptive depending on how they are utilized during the experiment. If the experimenter fills a survey during the experiment asking the subject to rate workload, situation awareness and fatigue, this can be intrusive and undesirable. Such intrusions may alter the sterile experimental environment which has been set up to reflect the real circumstances a pilot experiences during an actual flight. These interruptions may also affect the target measures as the conversations with the experimenters may alter the workload, situation awareness, response times and possibly fatigue levels. In order to minimize the effects of intrusive questioning, the pilots took post-run surveys to rate workload, situation awareness and fatigue at different stages of flight.

Data Tagging Tools

The lead experimenter was responsible for entering the time markers at desired points during the data collection and segmenting the data based on the stage of flight. These tasks were achieved using two special applications developed at the Operator Performance Laboratory, named the Phase Tagger and Cart GUI.

Phase Tagger was specifically used to tag the end of ATC calls and the beginning of pilot’s responses to ATC calls. The difference between these times was used to calculate the reaction times. It was also used to tag the times when the pilot closed his/her eyes due to fatigue. Figure 8 shows a screen shot of the Phase Tagger.

The Cart Recorder GUI is another tool that is used to segment the data. It contains a list of the phases of flight that have been described earlier. For every segment that is selected, it creates a new folder with text files that contain the raw data that can be used later for analysis. Figure 9 shows a screen shot of the Cart Recorder GUI. Figure 10 shows the lead experimenter's computer where Cart Recorder GUI and Phase Tagger are used. Table 4 lists the tags, their purpose and the tool that was used for the tagging task.

Table 4. List of Tags Used During the Data Collection

Tag Name	Purpose of Tag	Tag Application
Not Talking	Identifies the times when pilot is not talking to ATC	Phase Tagger
RT to ATC Call	Identifies the end of a call initiated by the pilot to ATC	Phase Tagger
Pilot Responds to ATC	Identifies the time when pilot starts to respond to ATC call	Phase Tagger
Eyes Closed	Identifies the times when the pilot's eyes are closed	Phase Tagger
Taxi and Takeoff	Identifies the duration of taxi and takeoff	Cart GUI
Climb to 19000	Identifies the time between takeoff and reaching 19000 feet	Cart GUI
Climb to 29000	Identifies the time between 19000 feet and 29000 feet	Cart GUI
Climb to 37000	Identifies the time between 29000 feet and 37000 feet	Cart GUI
Level Flight/Seattle Center	Identifies the phase when aircraft is at level flight at Seattle Center	Cart GUI
Level Flight/Salt Lake Center	Identifies the phase when aircraft is at level flight at Salt Lake Center	Cart GUI
Level Flight/Minn. Center	Identifies the phase when aircraft is at level flight at Minneapolis Center	Cart GUI
Level Flight/Chicago Center	Identifies the phase when aircraft is at level flight at Chicago Center	Cart GUI
Start Descent to 9000	Identifies the beginning of the descent to 9000	Cart GUI
Final Descent to Land	Identifies the phase when ATC clears the aircraft to land	Cart GUI
Touchdown and Taxi to Gate	Identifies the phase between touchdown and exit from runway	Cart GUI

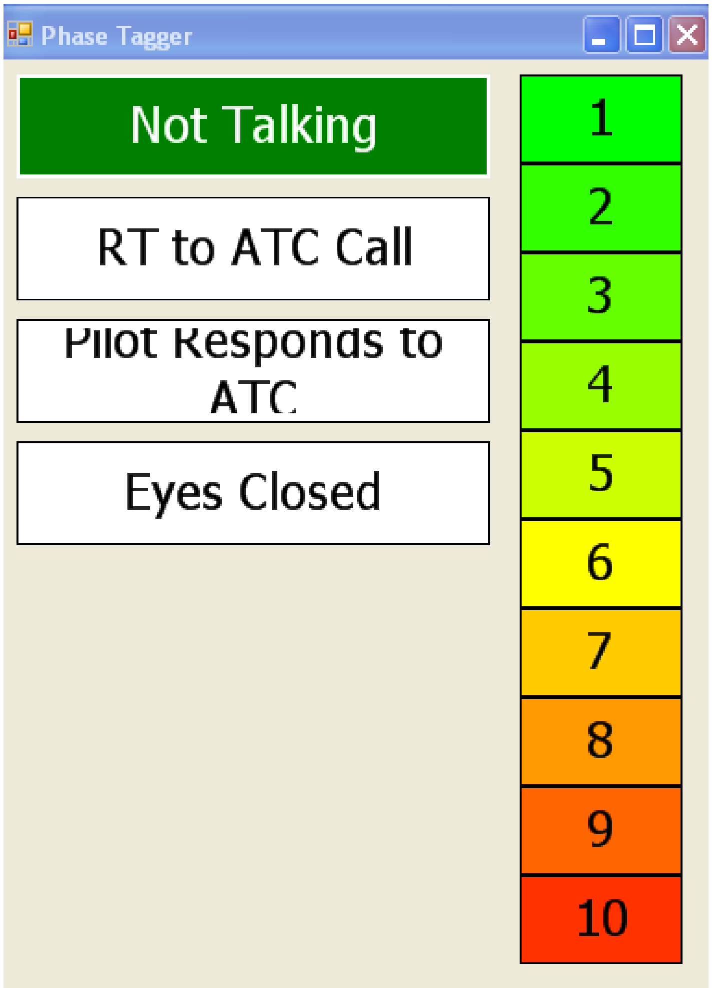
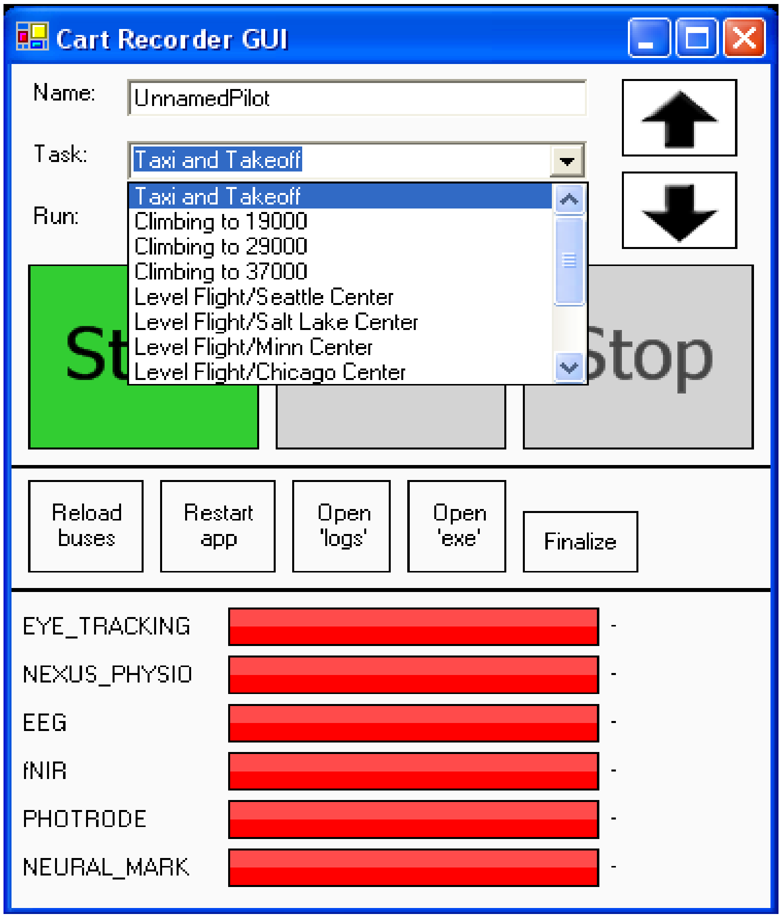


Figure 8. Phase Tagger



Name: UnnamedPilot

Task: Taxi and Takeoff

Figure 9. Cart Recorder GUI

Figure 10. Lead Experimenter's Station with the Data Tagging Tools

Detailed Flight Scenario

In this section, a typical flight scenario that the subjects went through is outlined. After the pilot was briefed about the experiment, wired up with the sensors and ready to start the simulation, the experiment commenced. A knee pad with airport drawings for Seattle Tacoma International and Chicago O’Hare International airports, the departure diagrams from Seattle, all the approach plates into Chicago and a 737 checklist were provided to the pilot.

Pilots were asked to arrive at 5:30pm. When the orientation and preparations were finished, it was around 7:00pm. Therefore, the simulation was set up such that the flight began at 7:20pm on August 1st. This was done so that the time the pilot experiences during the simulation was relatively close to the actual time of day. Also, this time of day provided enough sunlight on the runway to see the taxiway markings. Weather settings were set to clear. The flight management system, which would direct the autopilot when it was turned on, was already programmed to fly the route from Seattle to Chicago.

Each simulation started at the Seattle Tacoma International Airport. The experimenters made sure that the pilot's had their headsets on and the audio files were audible. The pilot was then instructed to make the initial call to Seattle Clearance, letting them know that they are ready to fly to Chicago O'Hare International Airport. The pilots were also briefed on how and when to turn on the autopilot so that the aircraft would follow the programmed flight path.

After the pilot's call, the experimenters started playing the ATC calls in sequence. At this point the experimenter started to play the ATC audio files from one of the laptops designated for this task. The experimenter had an Excel sheet that had instructions on when to play the audio files. The distance from waypoints and the time since takeoff were used to help the experimenter determine play times. A sample from this instruction sheet is shown in Table 5.

During each of the simulations, the first call came from Seattle Clearance which gave initial altitude instructions and transferred the pilot to Seattle Ground. The pilot acknowledged the request and contacted Seattle Ground for instructions.

Table 5. Air Traffic Control Instructions

Frequency	Station	Call	Distance	Time
128	Seattle Clearance	IFR Clearance to Chicago	When ready	
128	Seattle Clearance	Readback correct, handoff to Ground	After read-back	
121.7	Seattle Ground	Taxi to and hold short 16L	After request	
119.9	Seattle Tower	Position and Hold	After request	
119.9	Seattle Tower	Cleared for takeoff	Once in position	
119.9	Seattle Tower	Handoff to Departure on 120.1	Once airborne	1:00
120.1	Seattle Departure	Roger, own nav, climb to 9K	After check-in	1:30
120.1	Seattle Departure	Climb to 19K	Close to 9K	
120.1	Seattle Departure	Contact Center, 134.950	Passing through 10K	
134.95	Seattle Center	Roger		
134.95	Seattle Center	Contact Center, 120.300	47 to BLUIT	8:00
120.3	Seattle Center	Roger		
120.3	Seattle Center	Climb to 29K	When leveling at 19K	
120.3	Seattle Center	Contact Center, 132.6	20 to BLUIT	10:00
132.6	Seattle Center	Roger		
132.6	Seattle Center	Climb to 37K	When leveling at 29K	
132.6	Seattle Center	Contact Center, 126.100	36 to MWH	13:00
126.1	Seattle Center	Roger		
126.1	Seattle Center	Contact Center, 119.225	23 to ODESS	22:00
119.225	Seattle Center	Roger		
119.225	Seattle Center	Contact Center, 123.950	231 to HLN	27:00:00
123.95	Seattle Center	Roger		
123.95	Seattle Center	Contact Salt Lake Center, 133.400	130 to HLN	39:00:00
133.4	Salt Lake Center	Roger		
133.4	Salt Lake Center	Contact Salt Lake Center, 132.400	61 to HLN	46:00:00
132.4	Salt Lake Center	Roger		
132.4	Salt Lake Center	Contact Salt Lake Center, 133.400	30 to HLN	52:00:00
133.4	Salt Lake Center	Roger		
133.4	Salt Lake Center	Contact Salt Lake Center, 132.400	121 to BIL	60:00:00

The pilot tuned the radio to Seattle Tower frequency and taxied the plane to the designated runway. When ready, the pilot asked for take-off clearance and waited for take-off clearance. When cleared for take-off, the pilot followed the instructions on the checklist for take-off.

After takeoff, Seattle Tower transferred the pilot to Seattle Approach. Seattle Approach gave the pilot heading and altitude instructions and then transferred the pilot to Seattle Departure. Seattle Departure transferred the pilot to Seattle Center and gave new altitude instructions. At 37000 feet, the pilot began level flight. It is hypothesized that shortly after this phase has begun, the pilots began to transition from a high level of engagement to a low level of engagement.

Within Seattle Center, there were multiple times when the pilot was requested to switch radio frequencies. As the aircraft exited the area Seattle Center is responsible for, the pilot was transferred to Salt Lake Center. Again, multiple frequency changes took place in Salt Lake Center until the pilot entered the airspace of Minneapolis Center. Similar requests were made by ATC while in Minneapolis Center. After exiting Minneapolis Center, the aircraft entered Chicago Center. The aircraft was directed to O’Hare runway 22R via Janesville 5 approach.

Much of the cruise in Salt Lake and Minnesota Centers was mundane with not very many ATC calls. Also, the autopilot was engaged so the pilot was only required to make minor interventions to the aircraft state when necessary. This would mean that workload would be very low, which would translate into low level of pilot engagement.

CHAPTER 4

DATA ANALYSIS

Data was collected from 15 subjects over the course of six months. The data that was collected includes response times to Air Traffic Control calls and responses to a post-experiment survey. More specifically, the survey was used to collect the pilot’s subjective ratings on workload, situation awareness and fatigue experienced during each phase of flight. The goal is to use the data that was collected during and after the experiments to determine if there are statistically significant differences between the takeoff/ascent phase, the level flight phase and the descent/landing phase.

Radio Call Response Time Analysis

The radio call response time data that was collected during the experiments was broken down into multiple phases. These included taxi and takeoff, ascent to 19000 feet, ascent to 29000 feet, ascent to 37000 feet, level flight at Seattle Center, level flight at Salt Lake Center, level flight at Minneapolis Center, level flight at Chicago Center, descent to 9000, final descent to land and landing/taxi-to-gate. The data was initially broken down into more than three phases because this allowed for more control during the experiments. After the data collection, the phases were consolidated into three main phases. The first phase, which is takeoff/ascent included all the phases from taxi and takeoff upto level flight at Seattle Center.

The level flight phase included Seattle, Salt Lake and Minneapolis Centers. Chicago Center was included in descent because the first ATC call that was played during this center instructed the pilot to begin descending.

The radio call response times that were collected from all experiments were graphed using cumulative charts. These charts were chosen instead of other options because they display the distribution of the data well in addition to giving a good visual analysis on how the phases and the subjects differed. Figure 11 displays the cumulative chart for response times per subject. It can be clearly observed that the data is skewed.

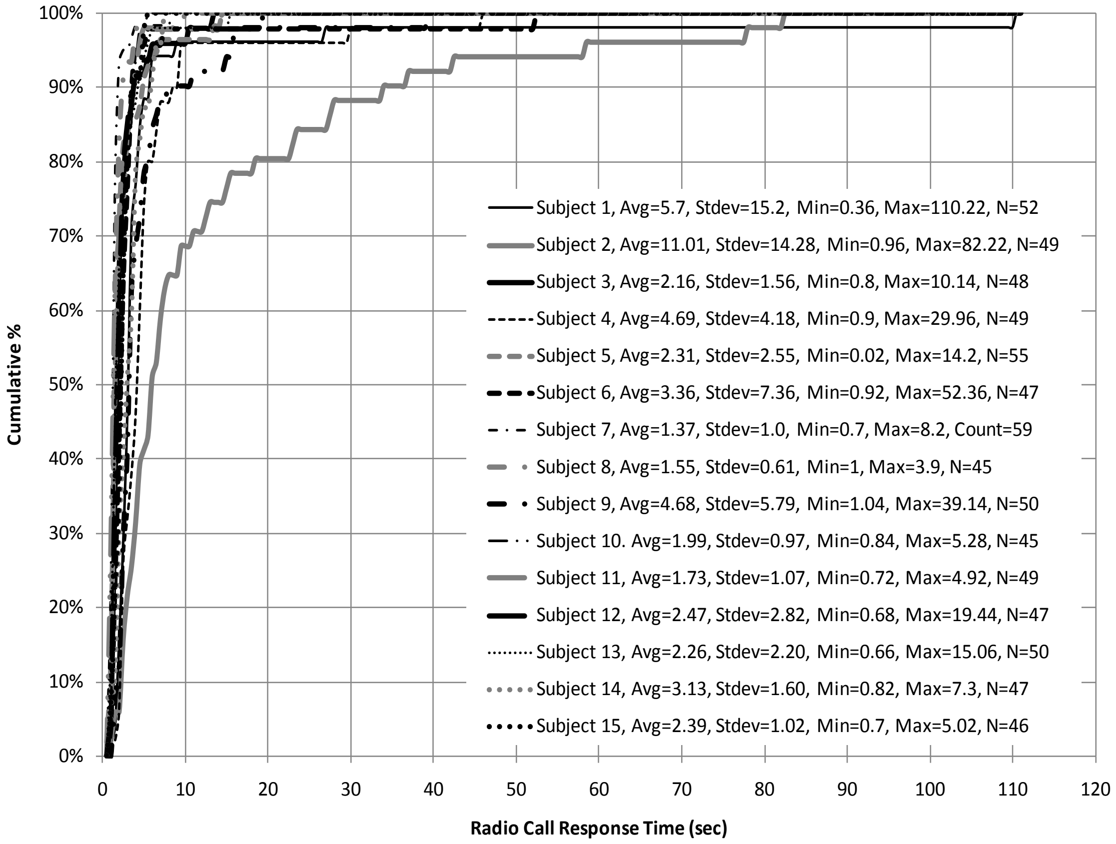


Figure 11. Cumulative Chart for Radio Response Times per Subject

This means that building any models or carrying out statistical tests to compare the different phases will be difficult using the original data because most statistical tests assume that the data are normally distributed (Fox, 2008, p. 54). A data transformation may be needed in order to conduct further tests using the radio call response times. The cumulative chart for response times per phase is presented in Figure 12.

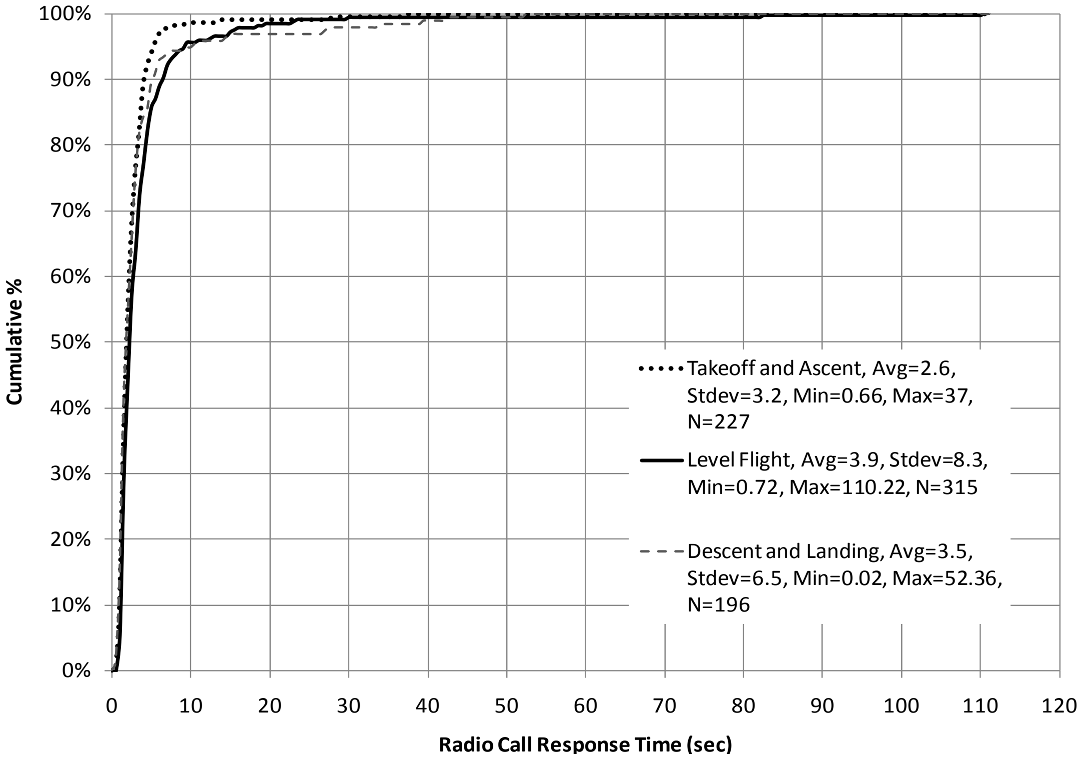


Figure 12. Cumulative Chart for Radio Response Times per Phase

It can be observed in both charts that most of the data points occur between zero and ten seconds. There are some data points that are well above 50 seconds. Table 6 lists which of these data points constitute the largest outliers that skew the data and identifies the reasons for their occurrence. It can be observed that the longest time occurred in the first experiment during level flight. The pilot must have been disengaged and did not notice the calls from Air Traffic Control to change frequency. The experimenters repeated the call twice with no response. The pilot eventually replied after the third try. The entire time span from the time the first call ended till the pilot replied on the third call was recorded as the reaction time to this particular call. The second longest response time was in the second experiment. The pilot responded to a call in Minneapolis Center (level flight) on the third try. Other outliers are in the range of 20 to 50 seconds. Some calls do not have reasons identified for their length. However, because they were under the outlier category, they were listed in the table along with the others.

Table 6. Table of Outliers

Subject	Phase	Reaction Time	ATC Center	Reason for Length
1	Level Flight	110.22	Salt Lake Center	Responded to ATC call on third try.
1	Descent/Landing	26.92	Chicago Center	Pilot was occupied reprogramming the CDU for runway change.
2	Takeoff/Climb	37	Seattle Departure	No reason identified.
2	Takeoff/Climb	27.98	Seattle Departure	No reason identified.
2	Level Flight	82.22	Minneapolis Center	Responded to ATC call on third try.
2	Level Flight	23.5	Minneapolis Center	Responded on 2nd call.
2	Level Flight	22.7	Minneapolis Center	Responded on 2nd call.
2	Level Flight	27.1	Chicago Center	Responded on 2nd call.
2	Descent/Landing	42.04	Chicago Tower	Answered call late due to hand flying place during second attempt to land.
2	Descent/Landing	33.58	Chicago Tower	Answered call late due to hand flying place during second attempt to land.
4	Level Flight	29.96	Salt Lake Center	No reason identified.
6	Level Flight	52.36	Chicago Center	No reason identified.
9	Level Flight	39.14	Chicago Center	No reason identified.

Since the normality assumption is an important requirement for many statistical tests, a transformation was conducted on the dataset to normalize the distribution of radio call response times. A log transformation was applied on the data. This was done because a log transformation is the preferred choice when the data is right skewed (Dehlert, 2000, p. 124). Figure 13 shows the distribution of the log of reaction times per subject.

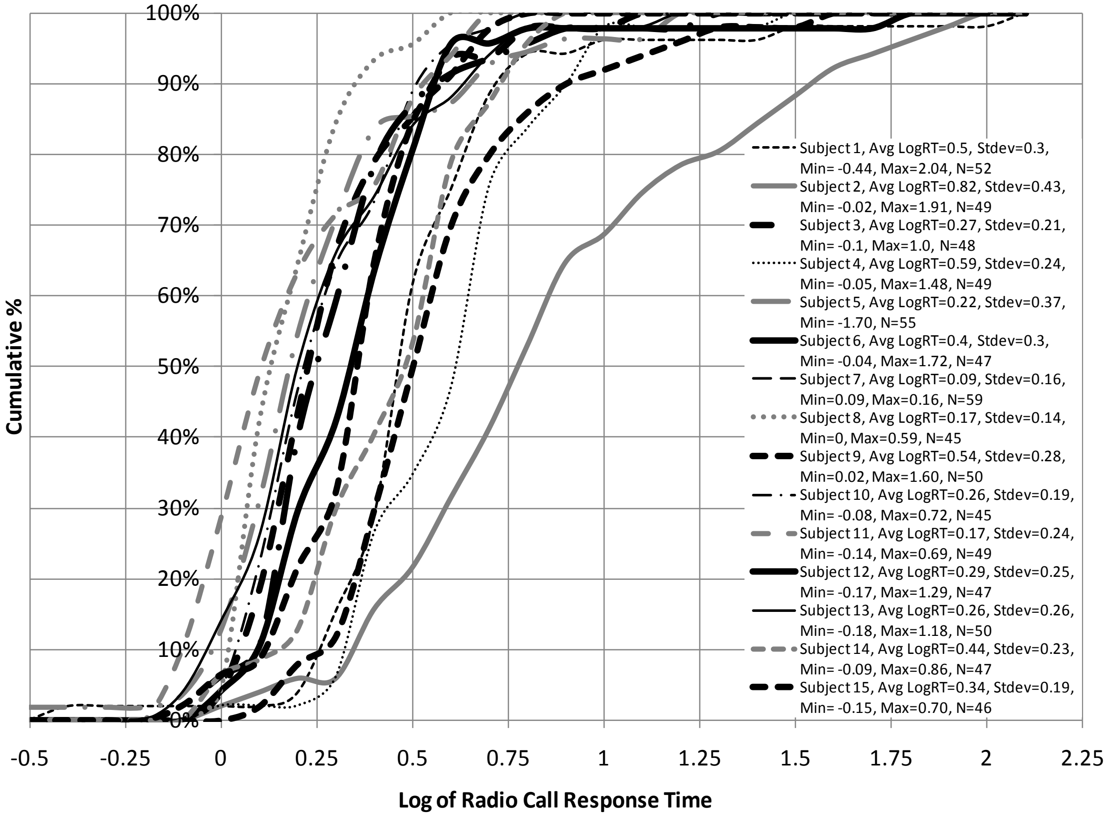


Figure 13. Cumulative Chart for the Log of Radio Call Response Times per Subject

It can be observed that the individual distributions for each subject are now normal. The transformed data was also graphed for each phase. This is shown in Figure 14.

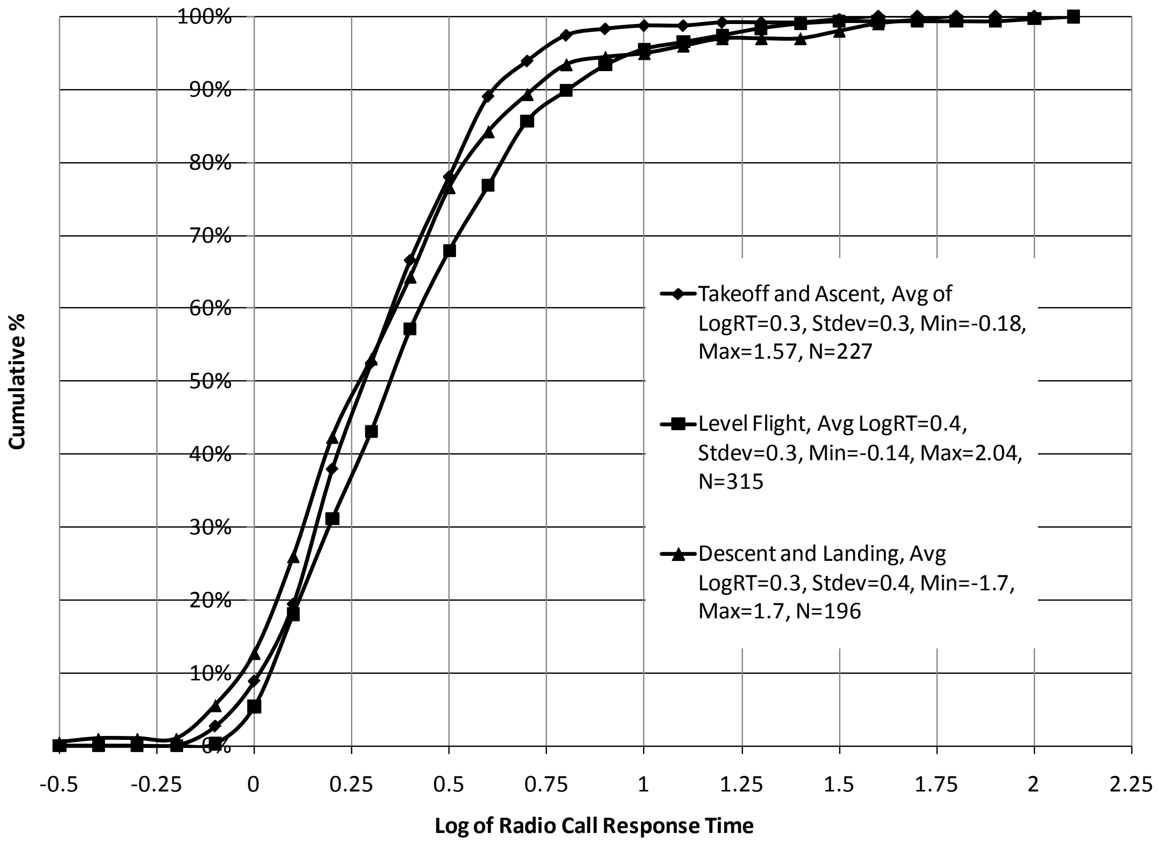


Figure 14. Cumulative Chart for Log of Radio Call Response Times per Phase

The cumulative chart shows that there may be a difference between subjects but the plot of the transformed data in Figure 14 does not show that there is an extreme difference between the phases. This was checked using Minitab. The transformed dataset was entered into Minitab and a general linear model was built using response time as the dependent variable and phase as an independent variable. A repeated measures analysis of variance was conducted on the data. Table 7 displays the repeated measures Anova table from the Minitab analysis.

It can be observed that, at the 90% confidence level, both phase and subject were significant contributors to the model built with response times as the dependent variable. Subject is also very significant at the 95% confidence level. Minitab was also used to conduct a pairwise analysis between the three phases to establish whether there was a difference between them and which ones were statistically significant.

Table 8 shows the results of this analysis. It can be observed that there is a statistically significant difference between takeoff/ascent and level flight where level flight has longer reaction times than those observed in the takeoff/ascent phase. As was expected, there is not a statistically significant difference between takeoff/ascent and descent/landing phases. During these phases of flight, pilots are more attentive as they are engaged in important preparatory tasks. There is not a significant difference between descent/landing and level flight at the 95% confidence level, but this difference is significant at the 90% confidence level. This indicates that if data was collected with more participants, the significance of the difference would be strengthened.

Table 7. Repeated Measures Anova Table for Log of Radio Call Response Times

Source	**DF**	**Seq SS**	**Adj SS**	**Adj MS**	**F**	**P**
Subject	14	26.14	23.22	1.66	10.64	0.000
Phase	2	0.86	0.8	0.4	2.54	0.096
Subject*Phase	28	4.42	4.42	0.16	2.39	0.000
Error	692	45.67	45.67	0.07		
Total	736	77.08				

Table 8. Pairwise Comparisons of Response Times between Phases

Phase Comparison	Difference of Means	SE of Difference	T-Value	Adjusted P-Value
Level Flight - Takeoff/Ascent	0.08	0.02	3.35	0.0023
Descent/Landing - Takeoff/Ascent	0.02	0.03	0.84	0.6812
Descent/Landing - Level Flight	-0.06	0.02	-2.31	0.0551

In addition to the analysis of response times presented thus far, a within subject analysis was conducted using t-tests on the transformed data. These tests present a more detailed look into which subjects contributed the most to the differences that were outlined in this section. The results of these t-tests can be found in Appendix B.

Survey Response Analysis

The survey was administered after the flight simulations were completed. Each question in the survey asked the subject to rate his workload, situation awareness and fatigue levels for each phase of flight. The survey, like the reaction times, spanned all 11 phases, which were described in the section above. The data that was collected was later consolidated into three phases, as was done with reaction times. It should be noted here that not all subjects finished the final phase of the simulation which was landing and taxi to gate. This was a short phase which only contained three ATC calls and one question in the survey, therefore the loss of this phase did not cause any issue with the analysis, especially when all the phases were consolidated into three phases. The complete survey is shown in Appendix A. Cumulative charts were generated using the data from the surveys, similar to the ones that were created for response times.

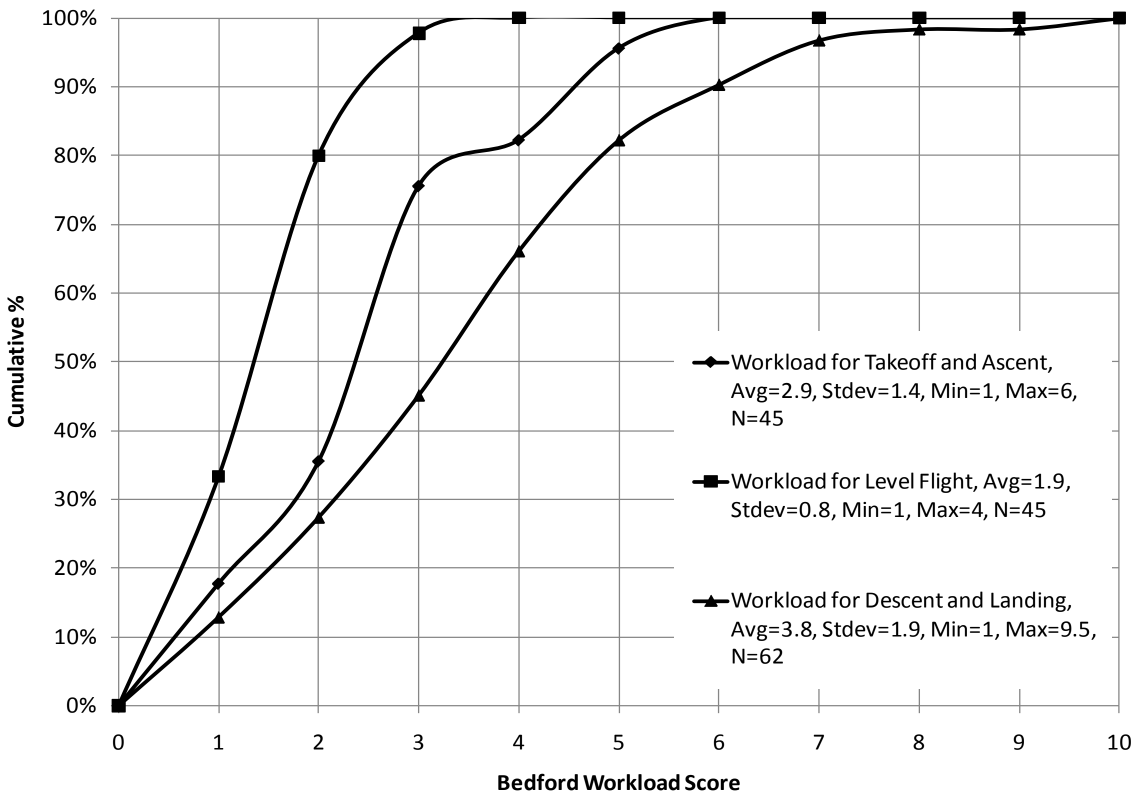


Figure 15. Cumulative Chart for Bedford Workload Scores

Figure 15 shows the cumulative chart for workload per phase. It can be observed here that there are differences between the three phases, especially between level flight and descent/landing phases. The graphical analysis of workload shows that descent/landing has the highest workload, which is followed by workload during takeoff/ascent and then level flight.

This makes intuitive sense as the pilots are more occupied with critical tasks ensuring that they land the airplane safely on the ground during the descent/landing phase, whereas during level flight, the tasks are more straightforward.

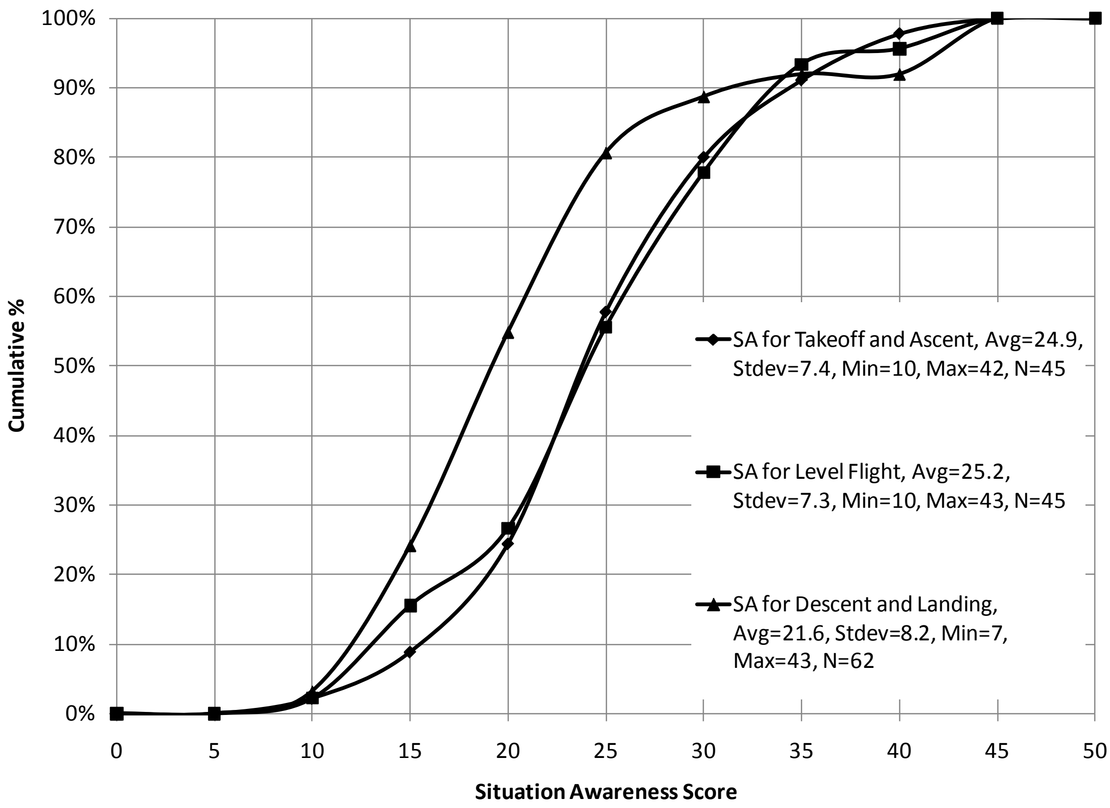


Figure 16. Cumulative Chart for Situation Awareness Scores per Phase

It can also be observed that the workload during takeoff/ascent is higher than that observed during level flight due to similar reasons. Whether these differences are statistically significant will be established using a similar approach where pairwise comparisons will be generated using Minitab after the graphical analysis of situation awareness and fatigue. This is an expected outcome, as pilots will feel more tired as the flight progresses. The change in fatigue, however, is not progressive. Fatigue experienced during descent/landing does not seem to be greater than what is experienced in earlier stages of flight. Figure 17 shows the cumulative chart for fatigue.

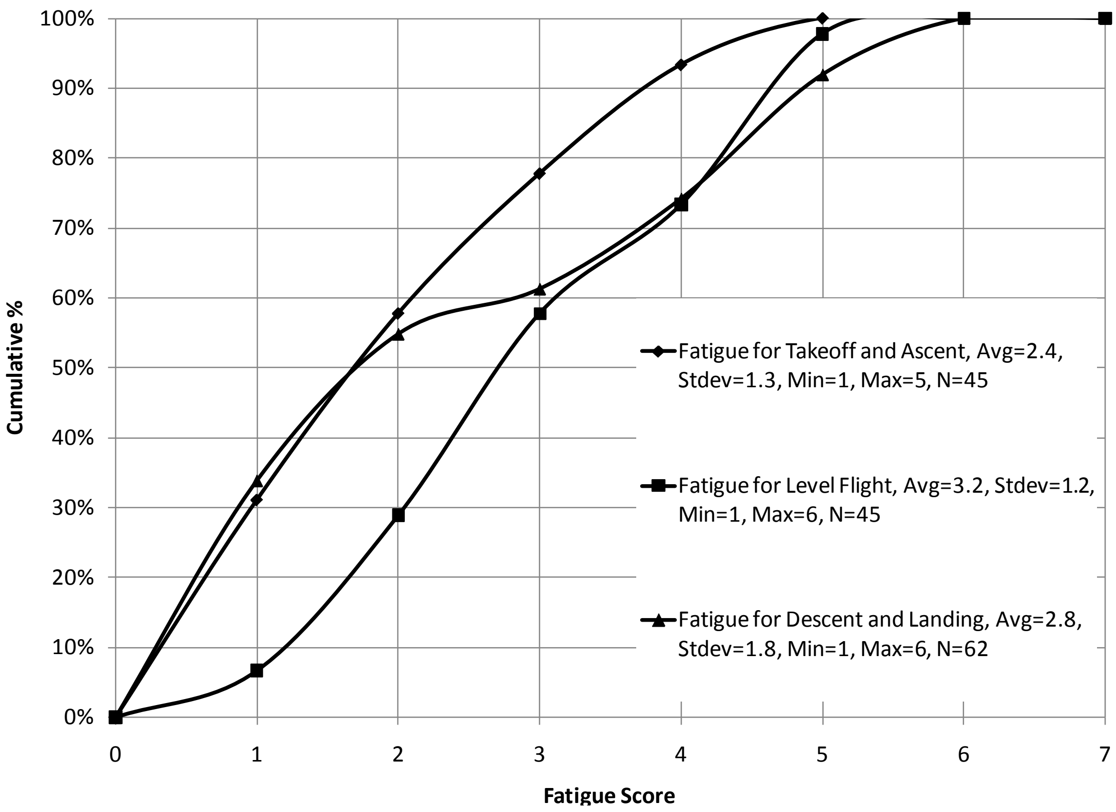


Figure 17. Cumulative Chart for Fatigue Scores per Phase

In this chart, it is observed that fatigue experienced during level flight is greater than what is felt during fatigue during takeoff/ascent.

As was done with the response times, a general linear model was built for each of the constructs that were scored with the surveys. Minitab was used to build the models. Subject and phase were the independent variables, where subject was identified as a random variable. Table 9 shows the repeated measures Anova table for the analysis done with workload. The p-values for phase and subject suggest that both are significant contributors to predicting workload. Table 10 shows the results of the pairwise comparisons between the phases of flight.

Table 9. Repeated Measures Anova Table for Workload

Source	**DF**	**Seq SS**	**Adj SS**	**Adj MS**	**F**	**P**
Subject	14	152.99	121.42	8.67	3.25	0.004
Phase	2	97.78	96.10	48.05	17.89	0.000
Subject*Phase	28	75.30	75.30	2.69	2.66	0.000
Error	111	112.17	112.17	1.01		
Total	155	438.24				

Table 10. Pairwise Comparisons of Workload between Phases

Phase Comparison	**Difference of Means**	**SE of Difference**	**T-Value**	**Adjusted P-Value**
Level Flight - Takeoff/Ascent	-1.04	0.25	-4.25	0.0001
Descent/Landing - Takeoff/Ascent	0.86	0.23	3.80	0.0006
Descent/Landing - Level Flight	1.91	0.23	8.42	0.0000

The adjusted p-values presented in Table 10 suggest that there is a significant difference in all of the comparisons. The difference of means suggest which of the phases have greater workload than others. The first comparison proves that there is a statistically significant difference between takeoff/ascent and level flight where workload experienced during takeoff/ascent is greater than what is experienced during level flight. The second comparison shows that pilots experienced greater workload during descent/landing compared to takeoff/ascent. Similarly, pilots dealt with greater workload during descent/landing than level flight. These results provide statistical proof that the tasks that pilots are engaged in during takeoff/ascent and descent/landing are heavier in terms of workload than the tasks that are experienced during level flight. We will now look at the results obtained from the statistical analysis of situation awareness and fatigue scores.

Table 11 shows the repeated measures Anova table for the model built with situation awareness scores as the dependent variable and phase and subject as the independent variables. The p-values show, as in the case of workload, that the variables phase and subject are significant contributors to the model with situation awareness.

Table 11. Repeated Measures Anova Table for SA Scores

Source	**DF**	**Seq SS**	**Adj SS**	**Adj MS**	**F**	**P**
Subject	14	6619.51	6215.66	443.98	8.85	0.000
Phase	2	466.27	467.55	233.77	4.61	0.019
Subject*Phase	28	1421.32	1421.32	50.76	5.9	0.000
Error	111	955.33	955.33	8.61		
Total	155	9462.44				

Table 12 shows the results of the pairwise comparisons between the phases with regard to the situation awareness scores. The comparisons show that pilots had greater situation awareness during the takeoff/ascent relative to the descent/landing phase. It is also observed from the results that pilots had better situation awareness during level flight compared to descent/landing. This is an odd finding because pilot's response times to ATC calls during level flight were found to be longer than those found in the takeoff/ascent phase. This may mean that despite the longer reaction times, pilots were still engaged even though they were not subjected to more challenging tasks like the ones experienced during takeoff/ascent or descent/landing phases. Finally, the comparison between level flight and takeoff/ascent in terms of situation awareness is not statistically significant, which was also observed in the graphical analysis.

Table 12. Pairwise Comparisons of SA between Phases

Phase Comparison	**Difference of Means**	**SE of Difference**	**T-Value**	**Adjusted P-Value**
Level Flight - Takeoff/Ascent	-0.47	0.87	-0.54	0.8540
Descent/Landing - Takeoff/Ascent	-3.72	0.80	-4.65	0.0000
Descent/Landing - Level Flight	-3.25	0.80	-4.07	0.0002

The final construct for which data was collected with the surveys is fatigue. Table 13 shows the repeated measures Anova table for the model built for fatigue as the dependent variable, where phase and subject were the independent variables. Table 14 shows the results of the pairwise comparison tests. The pairwise comparisons reveal that the difference in fatigue between level flight and takeoff/ascent is statistically significant. Also, it can be observed from the difference of means that level flight has a higher fatigue score compared to takeoff/ascent. This confirms the observation that was drawn from the graphical analysis earlier. It is also observed that there is a statistically significant difference between descent/landing and level flight. Level flight has the higher fatigue scores relative to descent/landing. The comparison between takeoff/ascent and descent/landing shows that there is no significant difference between these two phases.

These results confirm the observations that were made earlier in the graphical analysis section, where the change in fatigue was found to be non-progressive, and that fatigue does not progressively get worse with the progression of the flight.

Table 13. Repeated Measures Anova Table for Fatigue Scores

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Subject	14	240.44	216.63	15.47	5.81	0.000
Phase	2	20.91	20.88	10.44	3.88	0.033
Subject*Phase	28	75.53	75.53	2.70	9.52	0.000
Error	111	31.45	31.45	0.28		
Total	155	368.33				

Table 14. Pairwise Comparisons for Fatigue between Phases

Phase Comparison	Difference of Means	SE of Difference	T-Value	Adjusted P-Value
Level Flight - Takeoff/Ascent	0.96	0.18	5.17	0.000
Descent/Landing - Takeoff/Ascent	0.38	0.17	2.24	0.069
Descent/Landing - Level Flight	-0.58	0.17	-3.39	0.003

CHAPTER 5

CONCLUSION

In this study, a number of important questions were answered using a scientific approach. The intent of the study was to determine whether pilots experienced different engagement levels during a medium length commercial flight. The measures of choice were response times to Air Traffic Control calls and subjective measures such as workload, situation awareness and fatigue. The results of the study shed some light into what goes on with the workload and engagement levels of commercial pilots.

The simulations were set up as a night flight and subjects participated in the experiments in the evening to simulate the working conditions of commercial pilots whose responsibilities require flying more than once a day, including flying at night. At the onset of the study, it was expected that the results would show a clear difference between what pilots experienced during level flight and what they experienced during takeoff/ascent and descent/landing phases. This was based on what is already known about the tasks that the pilot is engaged in during each of the phases. Preparation for the flight, extensive communications with the airport tower and other tasks keep the pilots occupied, alert and focused during taxi, takeoff and ascent phases. During level flight, as the number of tasks thin out and the number of Air Traffic Control calls are few and far in between, pilots may become less engaged. The response time analysis showed that the time it takes the pilots to answer an Air Traffic Control call does indeed become longer during level flight than during the takeoff/ascent phase. Also, response times for level flight seemed to be longer than the response times during the descent/landing phase at the 90% confidence level.

The survey analysis also revealed some interesting results. One of the important findings was that fatigue, as reported by pilots, does not increase progressively. Pilots reported being more fatigued during level flight than the takeoff/ascent phase, but their fatigue did not continue to increase into the descent/landing phase. In terms of workload, pilots found the takeoff/ascent and descent/landing phases to be more challenging than level flight. This is consistent with the relatively large number of tasks that pilots are required to accomplish during takeoff and landing. It was also found that the descent/landing phase had greater workload than takeoff/ascent. This may be an indication that the descent and landing tasks are more involved than those during takeoff and ascent. The results of the analysis of situation awareness scores revealed that pilots had better situation awareness during takeoff/ascent than they did during descent/landing phases. However, this was not a very large difference which should lead to any alarming practical conclusions. It was also found that pilots had better situation awareness during level flight than they did during descent/landing. In the comparison between level flight and takeoff/ascent phases, the results indicate that pilots had better situation awareness during takeoff/ascent than they did during level flight. It should be noted, however, that this difference is not statistically significant. These results could lead to the argument that, in general, situation awareness got progressively lower during the flight. However, a close look at the differences of means suggest that these differences are not large enough to have any practical significance and do not lead to unsafe flying conditions.

The significant differences that were presented in this study, especially with regards to response times to Air Traffic Control calls indicate that this measure is suitable to determine when pilots may be experiencing low task engagement. Specific algorithms that utilize response times to measure pilot engagement can be designed as part of a comprehensive safety system that can alert Air Traffic Control when a pilot is identified to be in a state of low engagement. Air Traffic Controllers can then communicate with pilots to get them re-engaged with their primary tasks and take precautionary steps before dangerous situations occur.

The context of the experimental conditions should be considered when interpreting the results. The simulation was designed to reflect the conditions of an actual commercial flight as best as possible. However, simulations cannot be exactly the same as the actual situation that they are set up to represent. Therefore, it is noteworthy to mention some of the issues with the experimental setup in this study. Actual commercial flights have at least two pilots in the cockpit who are directly involved in flying the plane. Undoubtedly, the presence of a second pilot who assists with the tasks of the flight would reduce the workload on both pilots. What one pilot may miss could be noticed by the second pilot, so critical circumstances can be avoided. During takeoff, descent and landing, pilots are instructed to maintain a sterile cockpit where all conversations that take place between the pilot and co-pilot are related to flying the airplane. This means that there would not be other side conversations that could affect the pilot’s attention in such a way that their response to Air Traffic Control would be different than what was observed during this study. However, the effect the co-pilot would have during level flight might have lead to a more significant increase in the length of response times, especially if side conversations between the pilots took place. It would be difficult to understand how the different personalities of the subjects would influence the data that was collected. For instance, if two pilots who were both talkative were in the cockpit, there may have been a greater number of radio calls missed or responded to with a delay. Conversely, if two pilots who stayed away from unnecessary conversations participated in the same experiment, the data collected from them would have posed a challenge in having a fair comparison with the other experiments. It should also be noted that, during this study, pilots were asked to refrain from drinking any coffee or other drinks that has caffeine. This condition was applied because it would be difficult to determine how much caffeine one has taken before the experiment and how this might have affected their attention during the experiments. This may have had an effect on response times and fatigue. But, it could also be argued that under normal conditions, pilots are allowed to drink coffee and other beverages that may help them keep awake and be attentive during long flights. Therefore, this may have lead to shorter response times during level flight, which would have reduced the significance of the results. The only method that would lead to more concrete and reliable results for this type of a study would be to study the behavior of pilots under actual circumstances, in other words, during actual flights. However, the cost of modifying the cockpit, the cost of experimenters traveling to where airlines operate and other costs would make such a study extremely expensive. Furthermore, obtaining the permission of pilots and airlines alike would be very difficult, if not impossible. Therefore, a simulator study such as this one is the best alternative and the results can be validated with further studies with improved designs.

APPENDIX A: POST EXPERIMENT SURVEY

1. During takeoff and initial ascent to 19000 feet, how would you rate the following measures:
    a. Workload (*record a number based on the Bedford Workload decision tree*):
    b. Situation Awareness (*check a number for each construct*):

		Low						High
		1	2	3	4	5	6	7
DEMAND	Instability of Situation							
	Variability of Situation							
	Complexity of Situation							
SUPPLY	Arousal							
	Spare Mental Capacity							
	Concentration							
	Division of Attention							
UNDERSTANDING	Information Quantity							
	Information Quality							
	Familiarity							

    c. Fatigue (*circle a number from below*):

	Subject Fatigue
1	Fully Alert; Wide Awake; Extremely Peppy
2	Very Lively; Responsive, But Not at Peak
3	Okay; Somewhat Fresh
4	A Little Tired; Less Than Fresh
5	Moderately Tired; Let Down
6	Extremely Tired; Very Difficult to Concentrate
7	Completely Exhausted; Unable to Function Effectively; Ready to Drop

2. Between 19000 feet and 29000 feet, how would you rate the following measures:

    a. Workload (*record a number based on the Bedford Workload decision tree*):

    b. Situation Awareness (*check a number for each construct*):

		Low						High
		1	2	3	4	5	6	7
DEMAND	Instability of Situation							
	Variability of Situation							
	Complexity of Situation							
SUPPLY	Arousal							
	Spare Mental Capacity							
	Concentration							
	Division of Attention							
UNDERSTANDING	Information Quantity							
	Information Quality							
	Familiarity							

    c. Fatigue (*circle a number from below*):

	Subject Fatigue
1	Fully Alert; Wide Awake; Extremely Peppy
2	Very Lively; Responsive, But Not at Peak
3	Okay; Somewhat Fresh
4	A Little Tired; Less Than Fresh
5	Moderately Tired; Let Down
6	Extremely Tired; Very Difficult to Concentrate
7	Completely Exhausted; Unable to Function Effectively; Ready to Drop

3. Between 29000 feet and 37000 feet, how would you rate the following measures:

    a. Workload (*record a number based on the Bedford Workload decision tree*):

    b. Situation Awareness (*check a number for each construct*):

		Low						High
		1	2	3	4	5	6	7
DEMAND	Instability of Situation							
	Variability of Situation							
	Complexity of Situation							
SUPPLY	Arousal							
	Spare Mental Capacity							
	Concentration							
	Division of Attention							
UNDERSTANDING	Information Quantity							
	Information Quality							
	Familiarity							

    c. Fatigue (*circle a number from below*):

	Subject Fatigue
1	Fully Alert; Wide Awake; Extremely Peppy
2	Very Lively; Responsive, But Not at Peak
3	Okay; Somewhat Fresh
4	A Little Tired; Less Than Fresh
5	Moderately Tired; Let Down
6	Extremely Tired; Very Difficult to Concentrate
7	Completely Exhausted; Unable to Function Effectively; Ready to Drop

4. During level flight at Seattle Center, how would you rate the following measures:

    a. Workload (*record a number based on the Bedford Workload decision tree*):

    b. Situation Awareness (*check a number for each construct*):

		Low						High
		1	2	3	4	5	6	7
DEMAND	Instability of Situation							
	Variability of Situation							
	Complexity of Situation							
SUPPLY	Arousal							
	Spare Mental Capacity							
	Concentration							
	Division of Attention							
UNDERSTANDING	Information Quantity							
	Information Quality							
	Familiarity							

    c. Fatigue (*circle a number from below*):

	Subject Fatigue
1	Fully Alert; Wide Awake; Extremely Peppy
2	Very Lively; Responsive, But Not at Peak
3	Okay; Somewhat Fresh
4	A Little Tired; Less Than Fresh
5	Moderately Tired; Let Down
6	Extremely Tired; Very Difficult to Concentrate
7	Completely Exhausted; Unable to Function Effectively; Ready to Drop

5. During level flight at Seattle Center, how would you rate the following measures:

   a. Workload (*record a number based on the Bedford Workload decision tree*):

   b. Situation Awareness (*check a number for each construct*):

		Low						High
		1	2	3	4	5	6	7
DEMAND	Instability of Situation							
	Variability of Situation							
	Complexity of Situation							
SUPPLY	Arousal							
	Spare Mental Capacity							
	Concentration							
	Division of Attention							
UNDERSTANDING	Information Quantity							
	Information Quality							
	Familiarity							

   c. Fatigue (*circle a number from below*):

	Subject Fatigue
1	Fully Alert; Wide Awake; Extremely Peppy
2	Very Lively; Responsive, But Not at Peak
3	Okay; Somewhat Fresh
4	A Little Tired; Less Than Fresh
5	Moderately Tired; Let Down
6	Extremely Tired; Very Difficult to Concentrate
7	Completely Exhausted; Unable to Function Effectively; Ready to Drop

6. During level flight at Minneapolis Center, how would you rate the following measures:

   a. Workload (*record a number based on the Bedford Workload decision tree*):

   b. Situation Awareness (*check a number for each construct*):

		Low						High
		1	2	3	4	5	6	7
DEMAND	Instability of Situation							
	Variability of Situation							
	Complexity of Situation							
SUPPLY	Arousal							
	Spare Mental Capacity							
	Concentration							
	Division of Attention							
UNDERSTANDING	Information Quantity							
	Information Quality							
	Familiarity							

   c. Fatigue (*circle a number from below*):

	Subject Fatigue
1	Fully Alert; Wide Awake; Extremely Peppy
2	Very Lively; Responsive, But Not at Peak
3	Okay; Somewhat Fresh
4	A Little Tired; Less Than Fresh
5	Moderately Tired; Let Down
6	Extremely Tired; Very Difficult to Concentrate
7	Completely Exhausted; Unable to Function Effectively; Ready to Drop

7. During level flight at Chicago Center, how would you rate the following measures:

    a. Workload (*record a number based on the Bedford Workload decision tree*):

    b. Situation Awareness (*check a number for each construct*):

		Low						High
		1	2	3	4	5	6	7
DEMAND	Instability of Situation							
	Variability of Situation							
	Complexity of Situation							
SUPPLY	Arousal							
	Spare Mental Capacity							
	Concentration							
	Division of Attention							
UNDERSTANDING	Information Quantity							
	Information Quality							
	Familiarity							

    c. Fatigue (*circle a number from below*):

	Subject Fatigue
1	Fully Alert; Wide Awake; Extremely Peppy
2	Very Lively; Responsive, But Not at Peak
3	Okay; Somewhat Fresh
4	A Little Tired; Less Than Fresh
5	Moderately Tired; Let Down
6	Extremely Tired; Very Difficult to Concentrate
7	Completely Exhausted; Unable to Function Effectively; Ready to Drop

8. During the flight between Chicago Center and the initial call to begin descent to O'Hare, how would you rate the following measures:

    a. Workload (*record a number based on the Bedford Workload decision tree*):

    b. Situation Awareness (*check a number for each construct*):

		Low						High
		1	2	3	4	5	6	7
DEMAND	Instability of Situation							
	Variability of Situation							
	Complexity of Situation							
SUPPLY	Arousal							
	Spare Mental Capacity							
	Concentration							
	Division of Attention							
UNDERSTANDING	Information Quantity							
	Information Quality							
	Familiarity							

    c. Fatigue (*circle a number from below*):

	Subject Fatigue
1	Fully Alert; Wide Awake; Extremely Peppy
2	Very Lively; Responsive, But Not at Peak
3	Okay; Somewhat Fresh
4	A Little Tired; Less Than Fresh
5	Moderately Tired; Let Down
6	Extremely Tired; Very Difficult to Concentrate
7	Completely Exhausted; Unable to Function Effectively; Ready to Drop

9. Between the initial call to begin descent and leveling at 9000 feet, how would you rate the following measures:

    a. Workload (*record a number based on the Bedford Workload decision tree shown above*):

    b. Situation Awareness (*check a number for each construct*):

		Low						High
		1	2	3	4	5	6	7
DEMAND	Instability of Situation							
	Variability of Situation							
	Complexity of Situation							
SUPPLY	Arousal							
	Spare Mental Capacity							
	Concentration							
	Division of Attention							
UNDERSTANDING	Information Quantity							
	Information Quality							
	Familiarity							

    c. Fatigue (*circle a number from below*):

	Subject Fatigue
1	Fully Alert; Wide Awake; Extremely Peppy
2	Very Lively; Responsive, But Not at Peak
3	Okay; Somewhat Fresh
4	A Little Tired; Less Than Fresh
5	Moderately Tired; Let Down
6	Extremely Tired; Very Difficult to Concentrate
7	Completely Exhausted; Unable to Function Effectively; Ready to Drop

10. Between leveling at 9000 feet and the call transferring you to Chicago Tower, how would you rate the following measures:

    a. Workload (*record a number based on the Bedford Workload decision tree shown above*):

    b. Situation Awareness (*check a number for each construct*):

		Low						High
		1	2	3	4	5	6	7
DEMAND	Instability of Situation							
	Variability of Situation							
	Complexity of Situation							
SUPPLY	Arousal							
	Spare Mental Capacity							
	Concentration							
	Division of Attention							
UNDERSTANDING	Information Quantity							
	Information Quality							
	Familiarity							

    c. Fatigue (*circle a number from below*):

	Subject Fatigue
1	Fully Alert; Wide Awake; Extremely Peppy
2	Very Lively; Responsive, But Not at Peak
3	Okay; Somewhat Fresh
4	A Little Tired; Less Than Fresh
5	Moderately Tired; Let Down
6	Extremely Tired; Very Difficult to Concentrate
7	Completely Exhausted; Unable to Function Effectively; Ready to Drop

11. Between the call transferring you to Chicago Tower and arrival at your gate at O'Hare, how would you rate the following measures:

    a. Workload (*record a number based on the Bedford Workload decision tree*):

    b. Situation Awareness (*check a number for each construct*):

		Low						High
		1	2	3	4	5	6	7
DEMAND	Instability of Situation							
	Variability of Situation							
	Complexity of Situation							
SUPPLY	Arousal							
	Spare Mental Capacity							
	Concentration							
	Division of Attention							
UNDERSTANDING	Information Quantity							
	Information Quality							
	Familiarity							

    c. Fatigue (*circle a number from below*):

	Subject Fatigue
1	Fully Alert; Wide Awake; Extremely Peppy
2	Very Lively; Responsive, But Not at Peak
3	Okay; Somewhat Fresh
4	A Little Tired; Less Than Fresh
5	Moderately Tired; Let Down
6	Extremely Tired; Very Difficult to Concentrate
7	Completely Exhausted; Unable to Function Effectively; Ready to Drop

APPENDIX B: WITHIN SUBJECT ANALYSIS RESULTS

Table B1. Phase Comparison Results 1

		Regular T-Test		
Subject	**Comparison**	**P-value**	**Confidence Interval**	**Comments**
1	1 and 2	0.045	(-0.333, -0.004)	Significant difference: 2 > 1
	1 and 3	0.594	(-0.344, 0.205)	No significant difference
	2 and 3	0.487	(-0.193, 0.391)	No significant difference
2	1 and 2	0.005	(-0.670, -0.125)	Significant difference: 2 > 1
	1 and 3	0.330	(-0.528, 0.186)	No significant difference
	2 and 3	0.193	(-0.124, 0.577)	No significant difference
3	1 and 2	0.803	(-0.158, 0.123)	No significant difference
	1 and 3	0.369	(-0.317, 0.124)	No significant difference
	2 and 3	0.464	(-0.301, 0.143)	No significant difference
4	1 and 2	0.005	(-0.411, -0.080)	Significant difference: 2 > 1
	1 and 3	0.771	(-0.160, 0.120)	No significant difference
	2 and 3	0.013	(0.051, 0.4)	Significant difference: 2 > 3
5	1 and 2	0.283	(-0.265, 0.08)	No significant difference
	1 and 3	0.425	(-0.184, 0.424)	No significant difference
	2 and 3	0.141	(-0.076, 0.5)	No significant difference
6	1 and 2	0.000	(-0.316, -0.106)	Significant difference: 2 > 1
	1 and 3	0.079	(-0.533, 0.034)	No significant difference
	2 and 3	0.770	(-0.32, 0.242)	No significant difference
7	1 and 2	0.152	(-0.194, 0.031)	No significant difference
	1 and 3	0.863	(-0.088, 0.104)	No significant difference
	2 and 3	0.099	(-0.018, 0.197)	No significant difference
8	1 and 2	0.586	(-0.097, 0.169)	No significant difference
	1 and 3	0.568	(-0.098, 0.174)	No significant difference
	2 and 3	0.978	(-0.148, 0.151)	No significant difference

Table B2. Phase Comparison Results 2

		Regular T-Test		
Subject	**Comparison**	**P-value**	**Confidence Interval**	**Comments**
9	1 and 2	0.067	(-0.309, 0.011)	No significant difference
	1 and 3	0.477	(-0.336, 0.164)	No significant difference
	2 and 3	0.618	(-0.196, 0.321)	No significant difference
10	1 and 2	0.190	(-0.205, 0.042)	No significant difference
	1 and 3	0.027	(-0.327, -0.022)	Significant difference: 3 > 1
	2 and 3	0.208	(-0.243, 0.056)	No significant difference
11	1 and 2	0.000	(0.142, 0.387)	Significant difference: 1 > 2
	1 and 3	0.002	(0.131, 0.485)	Significant difference: 1 > 3
	2 and 3	0.586	(-0.122, 0.209)	No significant difference
12	1 and 2	0.746	(-0.158, 0.218)	No significant difference
	1 and 3	0.362	(-0.097, 0.255)	No significant difference
	2 and 3	0.578	(-0.128, 0.226)	No significant difference
13	1 and 2	0.440	(-0.097, 0.217)	No significant difference
	1 and 3	0.505	(-0.278, 0.14)	No significant difference
	2 and 3	0.175	(-0.319, 0.061)	No significant difference
14	1 and 2	8.461*10^-5	(-0.45, -0.183)	Significant difference: 2 > 1
	1 and 3	0.404	(-0.266, 0.111)	No significant difference
	2 and 3	0.007	(0.077, 0.401)	Significant difference: 2 > 3
15	1 and 2	0.639	(-0.206, 0.131)	No significant difference
	1 and 3	0.298	(-0.093, 0.288)	No significant difference
	2 and 3	0.033	(0.012, 0.259)	Significant difference: 2 > 3

REFERENCES

Brown, I. D. (1994). Driver Fatigue [Electronic Version]. *Human Factors, 36(2),* 298-314.

Caldwell, J. A. (2005). Fatigue in Aviation [Electronic Version]. *Travel Medicine and Infectious Disease, 3,* 85-96.

Cantin, V., Lavalliere, M., Simoneau, M., & Teasdale, N. (2009). Mental Workload When Driving in a Simulator: Effects of Age and Driving Complexity [Electronic Version]. *Accident Analysis and Prevention, 41,* 763-771.

Chen, S. (2009). *Pilot fatigue is like 'Having too much to drink'*. Retrieved November, 30, 2010, from http://articles.cnn.com/2009-05-15/travel/pilot.fatigue.buffalo.crash_1_pilot-fatigue-colgan-air-ntsb-web-site?_s=PM:TRAVEL

Costa, G. (2010). Fatigue and Biological Rhythms. Wise, J. A., Hopkin, V. D., & Garland, D. J. (Ed.), *Handbook of Aviation Human Factors* (pp. 11-1—11-23). Boca Raton, FL: CRC Press.

Dehlert, G. W. (2000). *A First Course in Design and Analysis of Experiments.* New York: W. H. Freeman and Company

Desmond, P. A., & Hancock, P. A. (2001). Active and Passive Fatigue States. Hancock, P. A., & Desmond, P. A. (Ed.), *Stress, Workload, and Fatigue* (pp. 455-465). Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Durso, F. T., & Alexander, A. L. (2010). Managing Workload, Performance, and Situation Awareness in Aviation Systems [Electronic Version]. *Human Factors in Aviation,* 217-242.

Endsley, M. R., Selcon, S. J., Hardiman, T. D., & Croft, D. G. (1998). A Comparative Analysis of SAGAT and SART for Evaluations of Situation Awareness [Electronic Version]. *Human Factors and Ergonomics Society Annual Meeting Proceedings, Aerospace Systems, 82-86.*

Endsley, M. R. (2000). Theoretical Underpinnings of Situation Awareness: A Critical Review. Endsley, M. R., & Garland, D. J. (Ed.), *Situation Awareness Analysis and Measurement* (pp. 3-32). Mahwah, NJ: Lawrence Erlbaum Associates, Inc. Publishers.

Endsley, M. R. (2010). Situation Awareness in Aviation Systems. Wise, J. A., Hopkin, V. D., & Garland, D. J. (Ed.), *Handbook of Aviation Human Factors* (pp. 12-1—12-18). Boca Raton, FL: CRC Press.

Fox, J. (2008). *Applied Regression Analysis and Generalized Linear Models.* Los Angeles: SAGE Publications, Inc.

Gawron, V. J. (2008). *Human Performance, Workload, and Situational Awareness Measures Handbook.* Boca Raton, FL: CRC Press

Harrison, Y., & Horne, J. A. (1999). One Night of Sleep Loss Impairs Innovative Thinking and Flexible Decision Making [Electronic Version]. *Organizational Behavior and Human Decision Processes, 78(2), 128-145.*

Harrison, Y., & Horne, J. A. (2000). The Impact of Sleep Deprivation on Decision Making: A Review [Electronic Version]. *Journal of Experimental Psychology: Applied, 6(3), 236-249.*

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research [Electronic Version]. *Advances in Human Psychology, 52, 139-183.*

Hitchcock, L., Bourgeois-Bougrine, S., & Cabon, P. (2010). Pilot Performance. Wise, J. A., Hopkin, V. D., & Garland, D. J. (Ed.), *Handbook of Aviation Human Factors* (pp. 14-1—14-19). Boca Raton, FL: CRC Press.

Job, R. F. S., & Dalziel, J. (2001). Defining Fatigue as a Condition of the Organism and Distinguishing It From Habituation, Adaptation, and Boredom. Hancock, P. A., & Desmond, P. A. (Ed.), *Stress, Workload, and Fatigue* (pp. 466-475). Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Jones, D. G. (2000). Subjective Measures of Situation Awareness. Endsley, M. R., & Garland, D. J. (Ed.), *Situation Awareness Analysis and Measurement* (pp. 3-32). Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Kantowitz, B. H., & Campbell, J. L. (1996). Pilot Workload and Flight Deck Automation. Parasuraman, R., Mouloua, M. (Ed.), *Automation and Human Performance* (pp. 117-136). Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Makishita, H., & Matsunaga, K. (2008). Differences of Driver's Reaction Times According to Age and Mental Workload [Electronic Version]. *Accident Analysis and Prevention*, *40,* 567-575.

Marks, A. (2009). *To reduce pilot fatigue, FAA moves to revise rules.* Retrieved November, 30, 2010, from http://www.csmonitor.com/USA/2009/0806/p02s20-usgn.html

Martinussen, M., & Hunter, D. R. (2010). *Aviation Psychology and Human Factors.* Boca Raton, FL: CRC Press

National Transportation Safety Board (n.d.). *Most Wanted Transportation Safety Improvements: Federal Issues.* Retrieved November, 30, 2010, from http://www.ntsb.gov/recs/mostwanted/aviation_reduce_acc_inc_humanfatig.htm

National Transportation Safety Board (2010). *Aircraft Accident Report: Loss of Control on Approach Colgan Air, Inc. Operating as Continental Connection Flight 3407 Bombardier DHC-8-400, N200WQ, Clarence Center, New York, February 12, 2009.*Washington, D.C.

Rehmann, J. T., Stein, E. S., & Rosenberg, B. L. (1983). Subjective Pilot Workload Assessment [Electronic Version]. *Human Factors, 25(3), 297-307.*

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 [Electronic Version]. *International Journal of Industrial Ergonomics, 39,* 490-500.

Ting, P., Hwang, J., Doong, J., & Jeng, M. (2008). Driver Fatigue and Highway Driving: A Simulator Study [Electronic Version]. *Physiology & Behavior, 94, 448-453.*

Van Dongen, H. P. A. (2006). Shift Work and Inter-Individual Differences in Sleep and Sleepiness [Electronic Version]. *Chronobiology International, 23(6),* 1139-1147.
Watson, A. R., Ntuen, C., & Park, E. (1996). Effects of Task Difficulty on Pilot Workload [Electronic Version]. *Computers ind. Engang, 31(1/2), 487-490.*