Spring 4-20-2020

# Evaluating Scenarios That Can Startle and Surprise Pilots

Rahim Daud Agha
*Embry-Riddle Aeronautical University*

# EVALUATING SCENARIOS THAT CAN STARTLE AND SURPRISE PILOTS

By

Rahim Daud Agha

A Thesis Submitted to the College of Aviation, School of Graduate Studies,
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Aeronautics

Embry-Riddle Aeronautical University
Daytona Beach, Florida
April 2020

EVALUATING SCENARIOS THAT CAN STARTLE AND SURPRISE PILOTS

By

Rahim Daud Agha

This Thesis was prepared under the direction of the candidate's Thesis Committee Chair, Dr. Jennifer E. Thropp Associate Professor, Daytona Beach Campus, and Thesis Committee Member Dr. Andrew R. Dattel, Assistant Professor, Daytona Beach Campus, and has been approved by the Thesis Committee. It was submitted to the College of Aviation, School of Graduate Studies in partial fulfillment of the requirements for the Degree of Master of Science in Aeronautics

Thesis Committee:

Jennifer E. Thropp
Jennifer E. Thropp (Apr 14, 2020)

Jennifer E. Thropp, Ph.D.
Committee Chair

Andy Dattel
Andy Dattel (Apr 15, 2020)

Andrew R. Dattel, Ph.D.
Committee Member

Donald S. Metscher
Donald S. Metscher (Apr 14, 2020)

Donald S. Metscher, D.B.A.
Program Coordinator
Master of Science in Aeronautics

Steven Hampton
Steven Hampton (Apr 17, 2020)

Steven Hampton, Ed.D.
Associate Dean, School of Graduate Studies,
College of Aviation

Alan J. Stolzer, Ph.D.
Dean, College of Aviation

Chris Grant

Christopher Grant, Ph.D.
Associate Vice President of Academics

04/20/2020
Date

Acknowledgments

I would take this opportunity to thank my committee chair, Dr Jennifer Thropp and committee member, Dr. Andrew Dattel, for providing support and guidance to conduct this study. I would also acknowledge the Office of Undergraduate Research Embry-Riddle Aeronautical University for funding this research.

I am thankful to Dr. Donald Metscher and Miss BeeBee Leong for their continuous support throughout my studies, to Dr. Scott Winter for being a mentor, to Dr Tom Haritos for guiding me through the initial stages of this research, and to Master of Science in Aeronautics graduate teaching assistants Sang-A, Vishnu, and Moh for providing timely feedback.

Abstract

Researcher: Rahim Daud Agha

Title: EVALUATING SCENARIOS THAT CAN STARTLE AND SURPRISE PILOTS

Institution: Embry-Riddle Aeronautical University

Degree: Master of Science in Aeronautics

Year: 2020

Startle and surprise on the flight deck is a contributing factor in multiple aviation accidents that have been recognized by multiple aviation safety boards. This study identified the effects startle and surprise had on commercial pilots with single and multi-engine ratings. Surprise is defined here as something unexpected (e.g., engine failure), while startle is the associated exaggerated effect of an unexpected condition (e.g., thunder sound). Forty pilots were tested in a basic aviation training device configured to a Cessna 172 (single-engine) and a Baron 58 (multi-engine). Each pilot flew the single- and multi-engine aircraft in a scenario that induced an uninformed surprise emergency condition, uninformed surprise and startle emergency condition, and an informed emergency condition. During each condition, heart and respiration rate, flight performance, and subjective workload measures were collected. The startle and surprise condition showed the highest heart and respiration rates for both aircraft. However, there was no difference in either the heart or respiration rates between the two aircraft for the informed condition. The subjective measures of mental, physical, and temporal demands, effort, and frustration were higher for the twin-engine aircraft when compared to the single-engine aircraft for all conditions. Performance (subjective) was not different between the single-and multi-engine aircraft for the surprise condition only. Objective flight performance, which was evaluated as a) participants' adherence to the engine failure checklist steps for single-engine aircraft; and b) altitude deviation for multi-engine aircraft, showed that pilots performed better in the informed emergency condition. Startle and surprise can be measured using heart and respiration rate as physiological markers, which can be used to evaluate if different flight simulator scenarios are startling, surprising, or neither. Potential applications of this study will help develop flight simulator scenarios for various unexpected conditions of different aircraft. Results of this study can potentially help pave the way for federal regulations that require training for startle and surprise.

*Keywords*: unexpected events, workload, vital signs, commercial pilots, training

Table of Contents

List of Tables

List of Figures

# Chapter I

## Introduction

The behavior of pilots in most Loss of Control-In Fight (LOC-1) accidents is unexplained (Federal Aviation Administration [FAA], 2017; National Transportation Safety Board [NTSB], 2010). In recent years after the crash of Air France 447 (AF447) and Colgan Air (CA3407), investigators have concluded that surprise and startle can disrupt the ability of most pilots to respond to an unexpected event. It is recommended to develop simulator scenarios that can surprise and startle pilots, and such scenarios to be used for training (FAA, 2017).

The terms *startle* and *surprise* are often used interchangeably in aviation operational practice (Rivera, Talone, Boesser, Jentsch, & Yeh, 2014). However, several authors have pointed out that startle and surprise are inherently different responses resulting from a different cause and effect (Martin, Murrary, Bates, & Lee, 2015; Rivera et al., 2014). The NTSB (2010) report on Colgan CA3407 and Bureau d'Enquêtes et d'Analyses (BEA; 2012) report on Air France (AF447) crash used "surprise" and "startle" as different terminologies.

Surprise is a cognitive-emotional response to an unexpected event that contradicts one's expectation, thus, forcing the person to change his or her understanding of the situation (Foster & Keane, 2015; Horstmann, 2006; Schützwohl, 1998; Niepel, Rudolph, & Schützwohl, 1998). Startle is a brief and highly physiological reaction to a sudden or threating stimulus, such as the sound of a gunshot (Martin et al., 2015; Thackray, 1988). According to Landman, Groen, van Paassen, Bronkhorst, and Mulder (2017a), eye blinks, contraction of the neck and facial muscles, the arrest of ongoing behavior, and report of anger or fear are measurable aspects of startle.

The cause and effect of surprise and startle also differ. For instance, a person can be surprised and startled at the same time, but the same person can also be surprised in the absence of a startle. Alternately, lightning strikes when flying through bad weather can be startling but not surprising. Even though studies indicate that surprises are common in the aviation industry, most of these events are not significant (de Boer & Hurts, 2017; Kochan, Breiter, & Jentsch, 2004). However, Landman et al. (2017a) suggested that surprise may impair the pilot's problem-solving abilities in some cases.

The crashes of AF447, CA3407, and Turkish Airlines Flight 1951 (TA1951) in 2009 were the main reason for the introduction of the term *startle* in aviation. In the aftermath of the CA3407 accident, the NTSB (2010) mentioned startle as one of the contributing causes of the accident. According to the NTSB (2010), the pilot was startled by the activation of the stick shaker, which led to confusion in the cockpit. Surprise has been a factor in airline accidents for more than three decades, having been mentioned in multiple accident reports. However, startle is a new concept in aviation, which was the center of debate after multiple crashes in 2009 (BEA, 2012; NTSB, 2010).

In a Notice of Proposed Rule Making (NPRM), the FAA (2014) proposed changes in the evaluation of certain training maneuvers in the Flight Simulation Training Devices (FTSD). It was presumed that training in realistic scenarios would help diminish the startle effect. Congruently, the FAA (2013) postulated that the goal of startle training is to provide pilots with startle experience, which, in turn, would allow for the effective recovery of the aircraft during emergencies. They ascertained that the physiological response of startle is difficult to re-create in a simulated training environment, while surprise can easily be created. They further suggested that surprise may lead up to startle if certain parameters are introduced at the right time. The FAA (2017) acknowledges that startle and surprise have been key factors in multiple Loss of Control In-Flight (LOC-I) accidents.

**Statement of the Problem**

According to Casner, Geven & Williams (2013), airline training events are highly scripted and predictable, which calls into question the abilities of pilots to recognize and respond to abnormal events. Similarly, BEA (2012) found that airline crews receive little training on how to manage a sudden and stressful event that requires quick and precise decision-making, while Casner et al. (2013) found that pilots are required by regulations to practice each abnormal event in training before the actual flight. Hence the pilots can anticipate that an unusual event will occur during the training.

Congruently, Moriarty (2015) found that pilots are accustomed to the idea that they will rarely face an emergency in their carrier, so to quickly transition from this presumed level of safety to a confusing (difficult) and life-threatening flight condition can startle and surprise pilots. While they are trained for emergencies, there are ample combinations of circumstances that pilots cannot train during simulator practice. According to Landman et al. (2017a), unexpected events induce a startling effect that may significantly impair performance. Situations that require human intervention usually are unanticipated, demanding quick and correct decision making.

Presently, no study has recorded the heart rate (HR) and respiration rate (RR) of pilots during events that can be surprising and startling. In addition, few studies have evaluated pilot performance during a startling event (Gillen, 2016; Martin et al., 2015), while few studies evaluated performance during events that were surprising (Kochan et al., 2004; Rankin, Woltjer & Field, 2016). However, in all these studies, there was no evidence to substantiate that the scenario was startling or surprising for the participants. Similarly, the workload during startle and surprise events in most cases will increase, but no study has evaluated workload during the aforementioned situations.

**Purpose Statement**

The purpose of this study was to assess the differences in the physiological response and performance between startle and surprise conditions. Heart rate and respiration rate were recorded to validate if the scenario was able to startle and surprise pilots. The study compared the scenarios for two aircraft (single-engine and multi-engine). The researcher also evaluated pilots' self-assessment of workload during startle and surprise events using the National Aeronautics and Space Administration Task Load Index (NASA-TLX).

**Significance of the Study**

The review and analysis of the Aviation Safety Reporting System (ASRS) safety database revealed that the term "startle" was mentioned in 134 coded reports, while the term "surprise" was mentioned in 904 coded reports (Rivera et al., 2014). The European Air Safety Agency (EASA) in Notice of Proposed Amendment (NPM) (2014) addressed the need for surprise and startle training. The training should address unexpected and stressful situations and prepare the crew to master these sudden events. However, for the training to be successful, there is a need to propose scenarios that can startle and surprise pilots. Presently, few studies have investigated if an unexpected event can startle and/or surprise pilots by recording vital signs (e.g., respiration rate) and skin conductance

(Landman et al., 2017a). This study is filling the research gaps by evaluating the effect of startle and surprise on pilot workload and vital signs (heart rate and respiration rate).

Surprise and startle training can help improve a pilot's performance during an emergency (EASA 2014; FAA, 2015). However, pilots cannot be trained for every possible emergency. For the training to be successful, flight training scenarios must be proposed that can startle and surprise the pilots. Most studies conducted on the topic only looked at a single type of aircraft. However, the researcher believes that for commercial pilots, type of aircraft can be a potential factor that can affect vital signs and performance.

**Research Question and Hypothesis**

The researcher for the current study investigated the following research questions. The study has 10 dependent variables and two independent variables (IV): aircraft with two levels, single-engine (Cessna 172SP) and multi-engine (Baron 58); and emergency scenario with three levels, uninformed surprise emergency, uninformed surprise and startle emergency, and informed emergency.

1. Do significant differences and interactions exist for heart rate (DV) and respiration rate (DV) based on the aircraft (IV) and emergency scenario (IV)?
2. Do significant differences and interactions exist for mental demand (DV), physical demand (DV), temporal demand (DV), subjective performance (DV), effort (DV), and frustration (DV) based on the aircraft (IV) and emergency scenario (IV)?
3. Do significant differences exist in the flight performance (DV) between the emergency scenarios (IV) for the multi-engine aircraft as measured by altitude

deviation?

4. Do significant differences exist in the flight performance (DV) between the emergency scenarios (IV) for the single-engine aircraft as measured by the number of engine-failure checklist steps followed?

The following null hypotheses were tested.

$H_0$1: There was no significant difference in heart rate between flying a single-engine and flying a multi-engine aircraft.

$H_0$2: There were no significant differences in heart rate among an uninformed surprise emergency, uninformed surprise and startle emergency, and informed emergency

$H_0$3: There were no significant interactions in the heart rate between the aircraft and the emergency scenario.

$H_0$4: There was no significant difference in the respiration rate between flying a single-engine and flying a multi-engine aircraft.

$H_0$5: There were no significant differences in respiration rate among an uninformed surprise emergency, uninformed surprise and startle emergency, and informed emergency.

$H_0$6: There were no significant interactions in the respiration rate between the aircraft and the emergency scenario.

$H_0$7: There was no significant difference in the mental demand between flying a single-engine and multi-engine aircraft.

$H_0$8: There were no significant differences in mental demand among an uninformed surprise emergency, uninformed surprise and startle emergency, and informed emergency.

$H_0$9: There were no significant interactions in mental demand between the aircraft and the emergency scenario.

$H_0$10: There was no significant difference in physical demand between flying a single-engine and multi-engine aircraft.

$H_0$11: There were no significant differences in physical demand between an uninformed surprise emergency, uninformed surprise and startle emergency, and informed emergency.

$H_0$12: There were no significant interactions in physical demand between the aircraft and the emergency scenario.

$H_0$13: There was no significant difference in temporal demand between flying a single-engine and multi-engine aircraft.

$H_0$14: There were no significant differences in temporal demand between an uninformed surprise emergency, uninformed surprise and startle emergency, and informed emergency

$H_0$15: There were no significant interactions in temporal demand between the aircraft and the emergency scenario.

$H_0$16: There was no significant difference in subjective performance between flying a single-engine and multi-engine aircraft.

$H_0$17: There were no significant differences in subjective performance between an uninformed surprise emergency, uninformed surprise and startle emergency, and informed emergency

$H_0$18: There were no significant interactions in subjective performance between the aircraft and the emergency scenario.

$H_0$19: There was no significant difference in effort between flying a single-engine and multi-engine aircraft.

$H_0$20: There were no significant differences in effort between an uninformed surprise emergency, uninformed surprise and startle emergency, and informed emergency.

$H_0$21: There were no significant interactions in effort between the aircraft and the emergency scenario.

$H_0$22: There was no significant difference in frustration between flying a single-engine and multi-engine aircraft.

$H_0$23: There were no significant differences in frustration between an uninformed surprise emergency, uninformed surprise and startle emergency, and informed emergency.

$H_0$24: There were no significant interactions in frustration between the aircraft and the emergency scenario.

$H_0$25: There were no significant differences in flight performance among the emergency scenarios for the multi-engine aircraft as measured by altitude deviation.

$H_0$26: There were no significant differences in flight performance among the emergency scenarios for the single-engine aircraft as measured by the number of engine-failure checklist steps followed.

**Delimitations**

The results of this study should be considered in the context of the following delimitations:

1. The study only evaluated two scenarios that have the potential to induce surprise and two scenarios that have the potential to induce startle and

surprise.

2. The study only observed commercial pilots with single-engine and multi-engine rating.
3. The study observed pilots on the X-Plane 11 Basic Aviation Training Device (BATD) while flying a Cessna 172 six-pack (C172SP) and Beechcraft Baron 58.
4. The study cannot represent the whole pilot population in the aviation industry.

**Limitations and Assumptions**

The researcher sought to mitigate potential limitations that can change the results. For this study, the following limitations were considered:

1. The BATD, though realistic in nature, involves simulation limits such as lack of motion and gravitational forces.
2. The participants were not randomly selected for this study.
3. The participant's behavior was observed under controlled conditions, which may not be the same in a natural environment.

The following assumptions were considered for this study:

1. The study assumed that each participant was physically fit and capable of flying a scenario in a BATD.
2. The participants have never received any training related to startle and surprise.
3. The study also assumed construct validity in that the surprise and startle emergency properly reflected (or induced) startle (and same for the surprise).

**Definitions of Terms**

| | |
|---|---|
| Aircraft | A device that is used or intended to be used for flight in the air (FAA and Aviation Supplies and Academics [ASA], 2015). |
| Startle | An uncontrollable, automatic muscle reflex, raised heart rate, and blood pressure elicited by exposure to a sudden, intense event that violates a pilot's expectations (FAA, 2015). |
| Surprise | An unexpected event that violates a pilot's expectations and can affect the mental processes used to respond to the event (FAA, 2015). |

**List of Acronyms**

| | |
|---|---|
| AC | Advisory Circular |
| AF447 | Air France 447 |
| AS | Automation Surprise |
| ASRS | Aviation Safety Reporting System |
| BATD | Basic Aviation Training Device |
| BEA | Bureau of Enquiry and Analysis for Civil Aviation Safety |
| CA3407 | Colgan Air 3407 |
| CAST | Commercial Aviation Safety Team |
| CRM | Crew Resource Management |
| DAB | Daytona Beach International Airport |
| DOT | Department of Transportation |
| DSB | Dutch Safety Board |
| ERAU | Embry-Riddle Aeronautical University |

| | |
|---|---|
| EASA | European Air Safety Agency |
| FAA | Federal Aviation Administration |
| HR | Heart Rate |
| HAI | Human Automation Interaction |
| IATA | International Air Transport Association |
| ILS | Instrument Landing System |
| LOC-I | Loss of Control In-Flight |
| NASA-TLX | National Aeronautics and Space Administration Task Load Index |
| NPA | Notice of Proposed Amendment |
| NPRM | Notice of Proposed Rule Making |
| NTSB | National Transportation Safety Board |
| PF | Pilot Flying |
| PNT | Pilot Not Flying |
| PARC | Performance-Based Aviation Rulemaking Committee |
| RR | Respiration Rate |
| SBT | Scenario-Based Training |
| SDT | Signal Detection Theory |
| SPSS | Statistical Package for the Social Sciences |
| TA1951 | Turkish Airlines 1951 |
| UPRT | Upset Prevention and Recovery Training |

# Chapter II

## Review of the Relevant Literature

A literature review was conducted to identify relevant peer-reviewed articles and studies. The articles reviewed were related to the supervisory role of humans in automation, signal detection theory (SDT), automation bias and complacency, startle effect, surprise, and LOC-I incidents. Further, this study also analyzed the investigation reports of NTSB, BEA, and the Dutch Safety Board (DSB). Similarly, the Advisory Circulars (ACs), NPRM, Notice of Proposed Amendment (NPA), and guidance material related to surprise, and startle was also studied. The review begins by discussing the role of humans in supervising automation. The literature review then discusses the SDT and relates it to automation bias and complacency. Further, startle and surprise are discussed in detail. Lastly, the review relates all the theory to actual airplane accidents in recent years and discusses pilot training. The purpose of this literature review is to give the readers a broad overview of the issue at hand and identify potential gaps in the literature that needs investigation.

### Human Supervisory Role in Automation

Casner et al. (2014) suggested that flight automation systems have now assumed the primary responsibility for many piloting tasks. Pilots are no longer required to carry out the control inputs manually. This paradigm shift is seeing the pilot's role being limited to automation monitoring or supervisory (Bhana, 2010). Supervision is defined as the activity of occasionally programming and receiving information from a computer, while the performer of the supervision is called a supervisor (Hew, 2016).

Boubin, Rusnock, and Bindewald (2017) ascertained that most automated systems cannot replace humans despite having superior performance. Evidently, human operators can provide value because they can make judgments in situations where automation fails. To accomplish most tasks safely, human and automation must become a team in which the human maintains a supervisory role, while the automation makes the majority of the decisions. However, Bhana (2010) found that instead of teaming up with automation, pilots normally over-rely on automation during unexpected events. This overreliance during supervisory duties can lead to improper evaluation of the signals and responses obtained during an automated process, hence decreasing the performance.

**Signal Detection Theory**

The processing of information in automated systems begins with the detection of an environmental event (Wickens, Hollands, Banbury, & Parasuraman, 2013). However, some stimuli are not identified, or they do not produce the anticipated response (Ponce, Polasko & Molina, 2016). As a result, the process does not accomplish the desired results. Hence, SDT is a tool that allows us to evaluate the signals and responses obtained during a process for increasing the performance of the complete process.

According to Wickens et al. (2013), SDT can be applied to the analysis of the detection performance by a human alone, by a machine, or by both humans and machines. The signal detection theory results in two responses: yes (signal is present) or no (signal not present). For example, there can only be two responses while assessing if the driving situation is hazardous (Wallis & Horswill, 2007). In this driving situation, there are two states of the world, and the observer is responsible for deciding which state has occurred (Wickens et al., 2013).

The combination of the states of the world (i.e., whether a stimulus such as a threat is either present or not present) and the operator's response to the stimulus (i.e., whether to treat the stimulus as a threat or not) produces the 2 x 2 matrix shown in Figure 1, generating four classes of joint events (Wickens et al., 2013). The joint events are labeled as misses, hits, false alarms, and correct rejections (Wallis & Horswill, 2007).

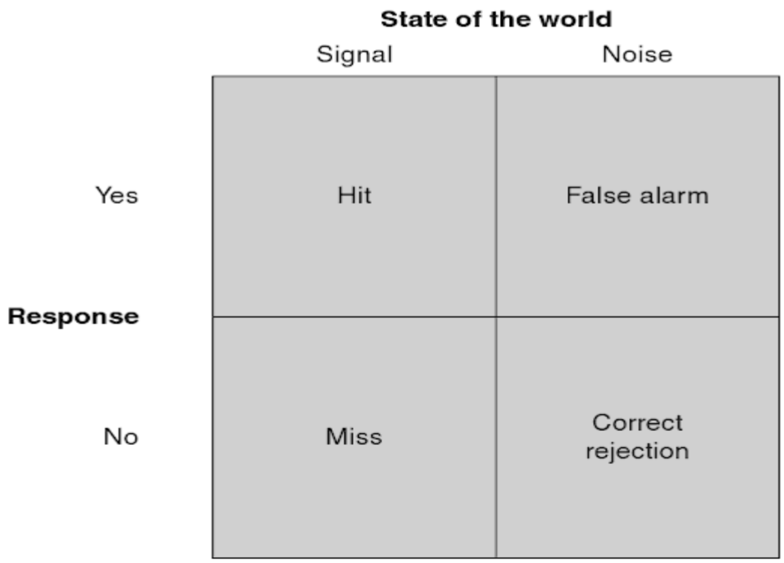


*Figure 1.* The four outcomes of SDT.

The Commercial Aviation Safety Team (CAST) and the Performance-Based Aviation Rulemaking Committee (PARC; 2013), suggested that pilots are reluctant to intervene during emergencies, as they rely too much on automated systems. The SDT can be applied to pilots, where the identification of the emergency is very important. For example, in the case of AF447, the aircraft was cruising when the pitot tube was chocked due to the formation of ice crystals (BEA, 2012). This led to false speed reading, which resulted in a stall warning (false alarm), the pilot flying (PF) did not recognize the false alarm. Eventually, severe turbulence, along with stall warning startled the PF who started pushing the aircraft nose up, further destabilizing the aircraft. However, the pilot not flying (PNF) recognized the false alarm but was not aware that the PF did not recognize the false alarm. The only thing the PF was supposed to do during this situation was just to fly the aircraft level for a while so that the ice crystals clear away, something he was trained for; however, his behavior was not consistent with training (BEA, 2012). In the case of CA3407, the automated system did pitch the aircraft nose down automatically when the stall was imminent (automation hit), but the PF pulled back on the control, causing the aircraft to stall and crash (NTSB, 2010).

**Automation Compliance and Reliance**

In the past two decades, the diverse effects of automation false alarms and misses on operator trust and performance have been scrupulously examined (e.g., Wickens et al., 2015; Parasuraman & Riley, 1997). Wickens, Clegg, Vieane, and Sebok (2015) suggested that imperfect automation in an alerting system can result in two types of automation errors. If the alerting system misses an actual dangerous event, then it can be an error of omission. Alternately, if the system issues a false alert, then it can be an error of commission. Each error will trigger a different Human Automation Interaction (HAI) behavior when the operator’s anticipation of automation failure is low. The 'overreliance' of an operator is associated with an error of omission, while the 'over-compliance' is associated with an error of commission.

According to Dixon, Wickens, and McCarley (2007), automation miss (not aware of changes made by automation) is more dangerous than an automation false alarm, as it may lead to true future alerts being disregarded. Automation miss is troublesome if the operator relies on the automation to identify all events (Wickens et al., 2015). Alternately, Dixon and Wickens (2006) suggested that an increase in automation miss rate reduces reliance, thus, causing the operator to monitor the raw data behind the automation miss, which diverts the attention of the operator from other concurrent tasks, causing a deterioration in performance. Congruently, Geels-Blair, Rice, and Schward (2013) found that automation misses tend to cause a reduction in reliance. Conversely, automation false alarms tend to cause a reduction in compliance.

According to Wickens et al. (2015), overreliance and over-compliance are closely associated with HAI problems of complacency and automation bias. Complacency is linked to the failure to be attentive in supervising automation before an automation miss. Similarly, Bhana (2010) suggested that in aviation, automation complacency is a term interchangeable with automation overconfidence and is described as pilots becoming complacent because they are overconfident and uncritical of automation. Parasuraman and Riley (1997) ascertained that automation complacency occurs when the operator excessively trusts the automation and fails to perform his supervisory duty. Lee (2008) suggests that automation seems to relieve people of tasks, but automation requires more attention to training and integration design. The BEA (2012) in the official report on AF447 accident, discovered that the autopilot disconnection made the crew aware that there was an issue. The crew was not sure why the autopilot disconnected. The autopilot disconnection led to a false instrument reading, which surprised both pilots and startled the co-pilot. The surprise and startle event started because the pilots were over-relying on automation during an emergency to assist with the task. Congruently, Parasuraman and Riley (1997) found that complacency has been a contributing factor in accidents in domains other than aviation. A widely cited example is the grounding of the cruise ship Royal Majesty, where it was found that the overreliance of the watch officers on automation systems was a probable cause of the accident. Alternately, Wickens et al. (2015) suggested that automation bias is the tendency to use automated cues as a replacement for information seeking. Eventually, this results in an error as the operator fails to notice a problem simply because the automation fails to detect them.

Similarly, Sauser, Chavaillaz, and Wastell (2016) found that automation bias is a tendency to follow the advice of the automation without deeming it correct or wrong. Parasuraman and Manzey (2010) ascertained that even with extended training and practice, the elimination of complacency and automation bias is difficult to achieve. Complacency and automation bias are closely linked with the pilots being startled and surprised during unexpected events (BEA, 2012; DSB, 2010).

**Startle**

Startle is a response to an unexpected stimulus that is common in all mammals (Simons, 1996). According to Koch (1999) and Martin et al. (2012), startle is an event that triggers (a) a spontaneous physiological reflex, and (b) a behavioral startle response. The spontaneous startle reflex is fast, and it involves eye blinks, contraction of muscles, head ducks, and shoulder squat up. All these actions are to prepare the body for protection against unexpected situations. Similarly, the startle response includes spontaneous muscle contractions like a startle reflex, but also other emotional and cognitive resources. The startle reflex can trigger through auditory, visual, or tactile stimuli. It begins within 100 milliseconds (ms) of a stimulus being sensed (Carlsen, Chua, Inglis, Sanderson, & Franks, 2008; Yeomans, Li, Scott, & Frankland, 2002). Startle reflex lasts less than one second for a mild response and in the range of 1 s to 1.5 s for a severe response (Rivera et al., 2014; Ekman, Friesen, & Simons, 1985). Fetcho and

Mclean (2010) suggested that one way to elicit a startle is by presenting a sudden and loud sound. For example, a sound level of 80-90 decibels (dB) and more can startle a person (Fetcho & Mclean, 2018).

**Fear-potentiated startle.** According to Davis (1992) and Martin et al. (2015), the effects of startle increase when it occurs in the presence of a threating stimulus. Such aggregative responses which are worse in magnitude and have a longer-lasting effect are described as fear-potentiated startle. Similarly, Martin, Murrary, Bates, and Lee (2016) found that the fear-potentiated startle ensues from the combination of startling stimulus and the perception of threat, causing a fully developed stress reaction. According to Gillen (2016) and Rivera et al. (2014), unexpected events on the flight deck, such as aerodynamic stalls trigger fear-potentiated startle. Research has shown that fear-potentiated startle results in a significant cognitive disruption in most people (Gillen, 2016; Thackray &Touchstone, 1970).

**Cognitive consequences due to startle.** Gillen (2016) found that the cognitive resources of humans are finite. The author suggested that pilots need to devote more cognitive resources when flying manually, due to system failures or other issues. Further, the study found that during such system failures, most pilots tend to lose the ability to mentally project where the airplane is in space in regard to altitude, airspeed, and configuration.

Thackray and Touchstone (1970) determined that the recovery of performance after a startling event is quick. After the introduction of a startling stimulus, the maximum disruption occurs for the first 5 s. While significantly less disruption after the second 5 s interval lasts for 30 s to 1 minute. So, it is evident that the major performance decline occurs within the first 5 s following a startle. Similarly, Martin et al. (2012) found that startle reaction may last between 0.3 s to 1.5 s, depending on the severity. Thackray and Touchstone (1983) found that the participants who were exposed to a sudden high potency aural alarm signal made more incorrect responses on a task. On the other hand, participants who received a low-intensity alarm signal made less incorrect responses. Startle disrupts cognitive processing and negatively influences an individual's decision making and problem-solving abilities.

Gillen (2016), discussing the multiple resource theory, suggested that task performance depends on the relation between two parameters: cognitive resources available and the complexity of the situation. Ippel (1987; as cited in Gillen, 2016) suggested that task performance can be satisfactory if the amount of resources consumed by the task is lower than or equal to the available amount of memory. Alternatively, task performance can decline if the amount of resources consumed is greater than the available memory. Similarly, Martin et al. (2012) found that the performance of pilots during a startling event declined when the cognitive loads exceeded the number of available resources.

**Startle from the aviation perspective.** Gillen (2016) found that startle has been well researched and documented over the past 60 years. The concept of startle was introduced in the aviation industry after the accidents of CA3407, AF447, and TA1951. Since these accidents, few studies have been conducted related to startle among pilots (Gillen, 2016; Landman et al., 2017a; Landman, Groen, van Paassen, Bronkhorst, & Mulder, 2017b; Martin, Murray, & Bates, 2012; Martin et al., 2015; Martin et al., 2016; Rivera et al., 2014). According to IATA (2015a), startle is defined as the initial shortlived, spontaneous physiological and cognitive reactions to an unexpected event that initiate the normal human stress response. Similarly, as stated earlier, according to FAA (2015), startle is defined as an uncontrollable muscle reflex, and high blood pressure due to sudden exposure to an intense event.

Martin et al. (2015) suggested that airline pilots are regularly vulnerable to startling stimuli. However, most of these stimuli are irrelevant, but some pilots are exposed to critical events. For example, in the case of AF447, the aircraft was flying through severe turbulence and had a blocked pitot tube, which was startling for the PF, though this type of situation is common when flying over the Atlantic and some pilots will not find it startling or surprising. Similarly, Rivera et al. (2014) suggested that pilots can experience a variety of stimuli on the flight deck that can trigger a startle reflex and response. However, the main concern during or after a startle is the appropriateness of the pilot's decision choice and execution. The BEA (2012) in the accident report of AF447 found that the PF was so much startled that he verbally agreed to the PNF pitch down instruction, while still actually pitching up. Hence the need for an appropriate decision about complex flight tasks during startle is of utmost importance.

The BEA (2012), in the investigative report of AF447, found that the poor management of the startle effect was one of the causes of the accident. Further, it was found that the startle effect played a significant role in the destabilization of the flight path. Interestingly it was found that the crew had performed training like the actual operating scenario, but the training did not consider the ramification of startle on their behavior. Similarly, the NTSB (2012), in the accident report of CA3407, found that the captain's behavior was consistent with startle and confusion. The PF, in the case of

CA3407, reacted similarly to the PF in the case of AF447, where he pulled back on the control after stall when he should have pulled down.

Gillen (2016) substantiated that training related to the startle effect can improve pilot performance. The study considered a total of 40 crews (80 individual pilots). The participants were briefed that they will fly a profile in the simulator. Twenty flight crews (40 individual pilots) received training on how to deal with a startle. The training session included briefing and simulator practice. During simulator flight, the events that led to the pilot's being surprised, Gillen considered them as startle events. It appears that there was discord over the definition of the terms surprise and startle. As stated earlier, startle and surprise have different responses and they cannot be used in the same terminology. The study only considered two scenarios; however, there are ample scenarios to research.

**Physiological measurements for startle.** As stated earlier, startle is associated with high heart rate and blood pressure (FAA, 2015). To date, no study has performed physiological measurements for pilots during a startling event. Landman et al. (2017a) ascertained that different factors need to be manipulated to induce a startle or surprise. A loud explosion or an abrupt sound can induce a startle. When compared to a surprising event, a startling event more likely will lead to higher blood pressure and heart rate. A startling event may also impact the respiration rate of pilots; however, there has been no study done to investigate those effects.

## Surprise

The definition of surprise varies and depends on the context it is being used. The basic definition of surprise as postulated by Bredehoeft (2005), is the gathering of new information that voids the original conceptual model. Similarly, surprise is also defined as a cognitive-emotional response to something unanticipated, which leads to a conflict between one's expectation and awareness of one's environment (Horstmann, 2006; Schützwohl, 1998). Rivera et al. (2014) suggested that the main concern with surprise is that it can interrupt an ongoing task. For example, Horstmann (2006) found that a surprising occurrence (unannounced visual event) hampered the continuous action of 78% of the participants who were doing rapid alternate finger tapping. In that study, the average duration of the interruption lasted about 1 s.

**Surprise in aviation perspective.** According to IATA (2015a), surprise is defined as the emotionally based acceptance of a difference in what is actual and what was expected. Similarly, as stated earlier, according to the FAA (2015), surprise is defined as an unexpected event that can influence the mental process of a pilot to respond to an emergency. Congruently, Hilscher, Breiter, and Kochan (2012) found that flight crew (pilots) can be surprised due to an unexpected event during the flight that contradicts their expectations. Casner et al. (2013) ascertained that pilots have considerable difficulties with applying learned procedures when they are surprised. Casner et al. showed that the time to respond to an event increases when this event comes abruptly. The results of the study suggested pilots to stop testing and practicing unexpected events, in the same way, every time because it becomes excepted and predictable.

**Automation surprise.** Present-day cockpit automation consists of auto-throttle, autopilot, flight management system, flight director, and multiple other systems (Moriarty, 2014). These systems link together to ensure partially or fully automated flight when required. Kochan et al. (2004) identified automation as one of the major culprits in triggering flight deck surprises.

According to Woods and Sarter (2002; as cited in de Boer and Hurts, 2017), three factors increase the probability of automation surprise: (a) acts by the automation without immediately preceding crew input, (b) gaps in the pilot's mental representation on automation, and (c) sub-standard feedback of the automation. Similarly, Dekker (2014) found that automation surprise is the outcome of an anomaly between one's expectation and the actual system behavior. This anomaly is only discovered when the crew notices an unexpected behavior on their part, which may already have led to serious consequences by that time.

De Boer and Hurts (2017) suggested that automation has increased in complexity, which has given rise to concerns. The behavior of pilots when dealing with automation is unexplained in approximately 46% of the aircraft accident reports and 60% of major aircraft incident reports (PARC/CAST, 2013). Congruently, EASA (2013) suggested that unanticipated events are one of the leading issues in-flight crew automation that requires attention. The Dutch Safety Board (DSB) (2010), in the accident report of TA1951, found that the crew was completely taken by surprise because they were over-relying on the automation, which had failed their expectations. The aircraft had a faulty altimeter that went unnoticed and led to a low altitude stall which crashed the aircraft. De Boer and Hurts (2017) found that without AS, the mismatch between reality and the pilot's expectation continues until reality readjusts with the expectation without them ever knowing.

**Physiological measurements for surprise.** Landman et al. (2017b) and Bruna, Levora, and Holub (2018) recorded physiological measurements while pilots were flying a surprise scenario in a flight simulator. Landman et al. (2017b) did not find a significant difference in heart rate between a surprise and a normal condition. However, they did find a significant difference in the skin conductance between the conditions. Bruna et al. (2018) found a significant difference in the respiration rate between different phases of flight when flying with and without navigation. The study also expected to see differences in the heart rate between the conditions, but those results are reported to another paper that is yet to be published.

**Loss of Control In-Flight**

According to IATA (2015b), LOC-I is defined as the loss of an aircraft while it is in flight. This also includes aerodynamics stalls and upsets following an aircraft system failure. Gillen (2016) suggested that people often inaccurately associate landing and takeoff stage as the area with the highest risks. Boeing (2019) found that LOC-I is the single largest category of fatalities from 2009 through 2018, accounting for 1181 fatalities from 13 accidents.

According to Gillen (2016), most of the LOC-I accidents were the outcome of an unexpected event at the beginning of the accident timeline. LOC-I can begin suddenly following an inappropriate decision by the pilots. Similarly, Landman et al. (2017b) suggested that several of these events have been associated with inappropriate responses of the flight crew, and it is commonly suspected that surprise and startle contribute to such inappropriate responses (NTSB, 2010; BEA, 2012; DSB, 2010). Landman et al. (2017a) found that several recent flight safety events involving LOC-I, the unexpectedness of the situation is thought to have induced startle complicating the crew's troubleshooting (Martin et al., 2016).

**Training and Unanticipated Events**

The accident reports of AF447, CA3407, and TA1951 ascertain that the repose of pilots to an unanticipated event was different from the way they were trained. A study conducted by Martin et al. (2014) found pilots that did not receive startle and surprise training performed inappropriately or ineffectively. Alternately, pilots that received startle and surprise training performed effectively. Similarly, Casner et al. (2012) ascertained that there were considerable delays in the responses of pilots following an unexpected event. Congruently, Gillen (2016) found that pilots receiving startle training performed better during unexpected scenarios than pilots not receiving any training.

Gillen (2016) suggested that the training of airline pilots has become predictable in the current regulatory environment. Presently, the training will transfer well to predictable situations like the tests and most likely will not help during an actual emergency (Landman, van Oorschot, van Paassen, Groen, Bronkhorst, & Mulder, 2018). However, the FAA (2013) suggested different training scenarios with the emphasis on different areas, including startle and surprise. The design of the training scenarios is to provide the crew with a startle experience, which allows for the effective recovery of the airplane during an emergency. Similarly, the FAA (2017) also suggested training scenarios with the emphasis on startle and surprise.

Further, it was suggested that the flight instructors should plan upset scenarios that are likely to result in a startle or surprise. Finally, for the training to be beneficial, it should cover unexpected events over a wide array of circumstances and operations parameters (Casner et al., 2012). Presently there is no mandatory training for startle and surprise. However, Landman et al. (2018) hope that training will become mandatory in the near future. A potential gap in the literature is that few studies have proposed flight simulator scenarios that can startle and surprise pilots.

**Workload Measurement using NASA-TLX**

The NASA-TLX is a widely used tool to measure workload. The NASA-TLX has six subjective scales that are represented on a single page and include: (a) mental demand, (b) physical demand, (c) temporal demand, (d) performance, (e) effort, and (f) frustration (NASA, 1986). Workload measured using NASA-TLX is self-evaluated and is limited to the perception of the participant regarding the task (Fernandes & Braarud, 2015). The complexity and environment of a task can influence the workload. Therefore, during startle and surprise emergencies, it is fair to expect the workload score to go up when compared to an informed emergency. To date, no study has evaluated the workload of the pilot during emergencies that can be startling and surprising. However, Suppiah (2019) evaluated the workload using NASA-TLX during unexpected scenarios where the pilots were using electronic and then paper charts. The unexpected scenarios in this study might be surprising; however, they were not considered as a surprise emergency.

**Summary**

This literature review intended to provide the readers with an extensive background on which to base the study. The literature discussed the role of humans as supervisors of automation. It is known that humans over-rely on automated systems, which is the basis for automation bias. Similarly, automation tends to miss information or issue false alarms. The processing of information in automated systems was discussed in detail with the help of SDT.

Further, the literature also discussed surprise and startle and related it to the aviation industry. The consequences of startle and automation surprise on the pilot's performance were then discussed while relating it to some recent LOC-I incidents. The focus of the literature review was on the lack of training that airline pilots received and how proposing realistic scenarios can help with startle and surprise training. Potential gaps found during the literature review included the lack of studies that proposed flight simulator scenarios that can startle and surprise pilots. Also, few studies did measure vital signs such as HR and RR but only during surprise events. Among the few studies, the focus of all was on commercial airline pilots. Though airline pilots are important, if the training scenarios are practiced while pilots finish their training before flying for commercial airlines, it can ensure pilots are proficient in multiple potential scenarios that can be startling and surprising. Lastly, no study assessed the pilot workload during startle and surprise events.

# Chapter III

# Methodology

The methodology presented in this chapter supports the procedures used to investigate scenarios that can induce startle and surprise in pilots. In the current study, participants were asked to fly six scenarios in a BATD during which their HR and RR were recorded. After each flight, participants completed the NASA-TLX and an open-ended question, see Appendix C. Before data collection, permission to conduct the research was applied to the Institutional Review Board at ERAU, which was granted. The IRB approval and informed consent form are shown in Appendices A and B, respectively. This chapter provided the readers with an overview of the methods used.

## Population/Sample

Forty (36 males and 4 females) commercial pilots with single-engine and multi-engine ratings participated in this study. The participants were selected using convenience sampling from a pool of commercial pilots at Embry-Riddle Aeronautical University (ERAU). Participants were recruited using email announcements and paper flyers. Each participant who completed the study was paid $20.

**Apparatus and materials.** Six flight scenarios were created on an Elite PI-135 BATD using X-Plane 11 software. The health vitals HR and RR were recorded using the Nexus 10 device (MindMedia Inc., NL) that sampled at 1024 Hz and transmitted data to the data-acquisition computer via a cable. This study used the Statistical Package for the Social Sciences (SPSS) for statistical analysis of the data.

**Design and procedures.** This study was an experimental within-subject design. There were two independent variables (aircraft and emergency scenario) and eight dependent variables. The aircraft variable had two levels and the emergency scenario variable had three levels making it a 2 x 3 within-subjects design. The IVs and DVs are shown in Figure 1. The purpose of the study was not disclosed to the participants, who were informed that they would fly an Instrument Landing System (ILS) approach at Daytona Beach International Airport (DAB). When the study concluded, each participant was debriefed about the purpose of the study.

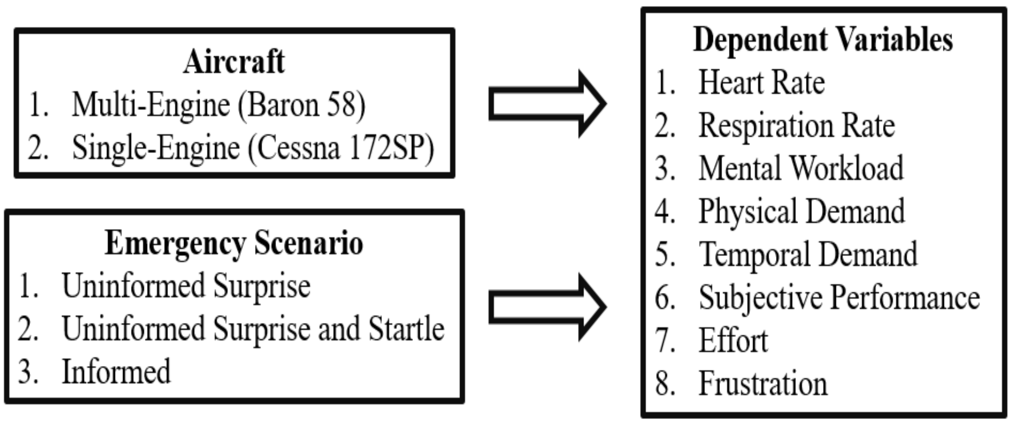


*Figure 2*. The independent and dependent variables for this study.

In the BATD, each participant flew six flight scenarios, where three configured as a C172SP and three configured as a Beechcraft Baron 58. Twenty participants flew three flights configured as a Baron aircraft followed by three flights in the Cessna; the other half flew Cessna, followed by the Baron. To induce startle, the researcher introduced loud bang and thunder sound during the simulation, where 20 participants heard a loud bang in the Baron and thunder noise in the Cessna, while the other half heard thunder sound in Baron and loud bang in Cessna. The sounds were played from a distance of 0.89 meters using a speaker that lasted 1 to 2 s, and a sound meter was used to measure the sounds in dB. The thunder sound was 85-89 dB, while the loud bang sound was 87-91 dB. Due to the unfamiliarity of most participants with the Baron, the researcher had them all fly a practice flight. The practice flight was a 3 nautical mile (nm) visual approach to 25R DAB. The researcher asked the pilots to maneuver the aircraft, increase and decrease speed using the throttle, climb and descent, and eventually land the aircraft.

After each flight, the researcher had the participant complete the NASA-TLX and an open-ended question. The NASA-TLX has six subjective scales which are rated on a scale of 1-20. A score of 1 is equivalent to the lowest possible workload, while a score of 20 is equivalent to the highest possible workload. An open-ended question "Which options best describes your perception of the task? (Check all that apply)" was used as a manipulation check where all participants were asked their perception about the three emergency scenarios for both aircraft, participants were asked if they found the emergency surprising, startling, both, or neither. The performance of the participants while flying was logged by the X-Plane software and evaluated differently for each aircraft. The description of the flight simulator scenarios for each aircraft is shown in Tables 1 and 2. The reason for having engine failure for Cessna at 1500 feet was to give the participant's time to initiate the checklist. For the multi-engine aircraft the engine failure was at 450 feet because the researcher had set up conditions for a missed approach (decision altitude = 234 feet) and having an engine failure at a high altitude would have resulted in some participants just flying the aircraft level to troubleshoot the problem, while some participants could have just aborted the landing. For the purpose of this study, the decision altitude used to evaluate the performance was changed to 180 feet as the participants were flying six-pack configuration.

Table 1

*Description of BATD Emergency Scenarios for Baron 58*

| Scenario | Scenario Parameters | Scenario Description |
|---|---|---|
| Uninformed Surprise Emergency | 3 nm ILS approach to 25R DAB | Engine failure at 450 feet with runway not visible at minimum |
| Uninformed Surprise and Startle Emergency | 3 nm ILS approach to 25R DAB | Engine failure at 450 feet with runway not visible at minimum. A loud bang or thunder sound at different altitudes |
| Informed Emergency | 3 nm ILS approach to 25R DAB | Engine failure at 450 feet with runway not visible at minimum, go missed |

*Note*. 20 participants heard the loud bang while 20 participants heard the thunder sound (with lightning simulated by turning the light on and off multiple times) in Baron uninformed surprise and startle emergency scenario.
Startle stimulus was introduced at 600, 500, 450, 400, 300 feet.

Table 2

*Description of Flight BATD Emergency Scenarios for* C172SP

| Scenario | Scenario Parameters | Scenario Description |
|---|---|---|
| Uninformed Surprise Emergency | 10 nm ILS approach to 25R DAB | Engine failure at 1500 feet with cloud layer set at 1000 feet |
| Uninformed Surprise and Startle Emergency | 10 nm ILS approach to 25R DAB | Engine failure at 1500 feet and engine fire at 1000 feet. A loud bang or thunder sound at different altitudes |
| Informed Emergency | 10 nm ILS approach to 25R DAB | Engine failure at 1500 feet with a cloud layer set at 1000 feet |

*Note*. 20 participants heard the loud bang while 20 participants heard the thunder sound (with lightning simulated by turning the light on and off multiple times) in Cessna uninformed surprise and startle emergency scenario.
Startle stimulus was introduced at 1700, 1500, 1400, 1300, 1000 feet.

## Sources of the Data

The HR and RR were measured at 32 samples per second using Nexus 10, which has been used in multiple published research studies. The performance of the participants for Cessna was measured by the number of checklist steps following an engine failure. The performance for Baron was measured in terms of deviation after crossing the decision altitude of 180 feet. Participants with an altitude of 180 or more deviated 0 feet from the decision altitude. The altitude was recorded by X-Plane software at five samples per second, which the researcher extracted after each flight. The workload and open-ended questions for each scenario were collected using a paper copy of the NASA-TLX.

**Instrument reliability.** The Elite-PI 135 BATD was successfully used in multiple published studies (Dattel et al., 2019; Suppiah, 2019) and is an FAA approved BATD. A pilot study was done using the BATD with a subject matter expert (SME) acting as the participant, who was satisfied with the functionality of the BATD and found the scenario parameters realistic. The SME was a full-time flight instructor with the ERAU flight department.

Bruna et al. (2018) measured the reparation rate of pilots while flying an emergency scenario using Nexus 10. Nexus 10 is being widely used in clinical research, and numerous studies have been conducted using this device (Bruna et al., 2018; Schuman & Killian, 2019).

**Instrument validity.** The NASA-TLX, Nexus 10, and the Elite-PI BATD are all valid instruments that are being used in different studies (Bruna et al., 2018; Dattel et al., 2019, & Suppiah, 2019). However, to ensure that all these devices were measuring what the researcher intended them to measure, the researcher manipulated both IVs. The order was counterbalanced by having half participants fly the Cessna followed by the Baron, and having the other half fly the Baron followed by the Cessna. For each aircraft, the first flight was an uninformed surprise emergency, the second flight was an uninformed startle and surprise emergency, and the third flight was informed emergency.

It was not possible to fully counterbalance the order of the emergency as that would have disclosed the emergency to the participant. However, the order of the two noise conditions was counterbalanced to counter an order effect due to the noise type where half the participants flying Cessna heard a loud bang while the other half heard thunder (lightning simulated by turning the light on and off). Conversely, half the participants flying the startle and surprise uninformed emergency in the Baron heard a loud bang while the other half heard thunder (lightning simulated by turning the light on and off).

**Treatment of the Data**

The HR and RR were recorded for all six flights using Nexus 10, and the output was saved as an Excel file where the researcher calculated the mean HR (beats per minute) and RR (breaths per minute), which the researcher entered in the main Excel data file. In the main data file, the researcher then entered the NASA-TLX self-assessed scores and the answer to the open-ended question for all six flights, asking participants about their perception task. The NASA-TLX scores were on a scale of 1 to 20, while the open-ended question had four possible answers (surprising, startling, both, or neither). The number of checklist steps followed after engine failure for the Cessna aircraft was reported on paper by the researcher and then entered in the main data file for each participant. For the Baron aircraft, the altitude at which the missed approach was initiated was entered, which the researcher extracted from X-plane software. X- Plane recorded the altitude in a text file that the researcher had to open, then visually see the altitude and enter it in the main data file for each participant. The data in the excel file was then open using the SPSS where the researcher then evaluated the data. All the data was rounded to 2 decimal places for consistency.

**Descriptive statistics.** The study presented the means and standard deviation for the HR, RR, and workload factors. Participant demographic and their answer to the open-ended question were also presented using pictorial representation.

**Hypothesis testing.** For the purpose of this research, the researcher evaluated each of the following 26 null hypotheses using a 2 x 3 with-subjects ANOVA. Post hoc tests were run for all significant interactions and significant main effect for emergency.

$H_0$1: There was no significant difference in heart rate between flying a single-engine and flying a multi-engine aircraft.

$H_0$2: There were no significant differences in heart rate among an uninformed surprise emergency, uninformed surprise and startle emergency, and informed emergency

$H_0$3: There were no significant interactions in heart rate between the aircraft and the emergency scenario.

$H_0$4: There was no significant difference in respiration rate between flying a single-engine and flying a multi-engine aircraft.

$H_0$5: There were no significant differences in respiration rate among an uninformed surprise emergency, uninformed surprise and startle emergency, and informed emergency

$H_0$6: There were no significant interactions in respiration rate between the aircraft and the emergency scenario.

$H_0$7: There was no significant difference in mental demand between flying a single-engine and multi-engine aircraft.

$H_0$8: There were no significant differences in mental demand among an uninformed surprise emergency, uninformed surprise and startle emergency, and informed emergency

$H_0$9: There were no significant interactions in mental demand between the aircraft and the emergency scenario.

$H_0$10: There was no significant difference in physical demand between flying a single-engine and multi-engine aircraft.

$H_0$11: There were no significant differences in physical demand between an uninformed surprise emergency, uninformed surprise and startle emergency, and informed emergency

$H_0$12: There were no significant interactions in physical demand between the aircraft and the emergency scenario.

$H_0$13: There was no significant difference in temporal demand between flying a single-engine and multi-engine aircraft.

$H_0$14: There were no significant differences in temporal demand between an uninformed surprise emergency, uninformed surprise and startle emergency, and informed emergency

$H_0$15: There were no significant interactions in temporal demand between the aircraft and the emergency scenario.

$H_0$16: There was no significant difference in subjective performance between flying a single-engine and multi-engine aircraft.

$H_0$17: There were no significant differences in subjective performance between an uninformed surprise emergency, uninformed surprise and startle emergency, and informed emergency

$H_0$18: There were no significant interactions in subjective performance between the aircraft and the emergency scenario.

$H_0$19: There was no significant difference in effort between flying a single-engine and multi-engine aircraft.

$H_0$20: There were no significant differences in effort between an uninformed surprise emergency, uninformed surprise and startle emergency, and informed emergency

$H_0$21: There were no significant interactions in effort between the aircraft and the emergency scenario.

$H_0$22: There was no significant difference in frustration between flying a single-engine and multi-engine aircraft.

$H_0$23: There were no significant differences in frustration between an uninformed surprise emergency, uninformed surprise and startle emergency, and informed emergency

$H_0$24: There were no significant interactions in frustration between the aircraft and the emergency scenario.

One-way repeated-measures ANOVAs were run to test the following null hypotheses. A Bonferroni post hoc was run if the ANOVA was significant.

$H_0$25: There were no significant differences in flight performance among the emergency scenarios for the multi-engine aircraft as measured by altitude deviation.

$H_0$26: There were no significant differences in flight performance among the emergency scenarios for the single-engine aircraft as measured by the number of engine-failure checklist steps followed.

**Qualitative analysis.** Participants ($N = 25$) that had the startle reflex or seemed surprised were asked if they wish to provide their narrative of the flight simulator scenarios and answer questions asked by the researcher. The researcher visually observed startle reflex which was associated with fast eye blinks, head ducks, or shoulder squat up. While surprise was associated with small gasps, sudden onset of anger, or a verbalization of what was being felt.

The questions were specific to the pilot's behavior while flying. For example, most participants were asked the reason for not initiating a go-around at the decision altitude for the multi-engine aircraft. Similarly, participants were asked the reason for not turning the fuel selector off during engine fire for the single-engine aircraft. Lastly, most participants were also asked why their performance was the best while flying the informed emergency condition. The researcher used the narrative of the pilots to interpret the results of the study.

# Chapter IV

# Results

This chapter presents the results for this study, which include descriptive and inferential statistics. Based on the results of null hypothesis testing, decisions to either retain or reject the null hypothesis were made (using a criteria of $\alpha = .05$). The assumptions for each statistical test were tested, and if any assumption were violated, appropriate measures were taken.

## Descriptive Statistics

Forty participants (male: $n = 36$, female: $n = 4$) participated in this study. All participants were asked about their perception of the emergency. For all three emergencies for both aircraft, participants were asked if they found the emergency surprising, startling, both, or neither. The results for participant perception are presented in Table 3.

Table 3

*Participant Perception of Emergency Scenario Based on Aircraft*

| Variable | Multi-Engine | | | | Single-Engine | | | |
|---|---|---|---|---|---|---|---|---|
| | Surprising | Startling | Both | Neither | Surprising | Startling | Both | Neither |
| Uninformed Surprise | 22 | 5 | 9 | 4 | 26 | 2 | 5 | 7 |
| Uninformed Surprise and Startle | 8 | 9 | 20 | 3 | 10 | 8 | 20 | 2 |
| Informed | 6 | 4 | 2 | 28 | 3 | 4 | 3 | 30 |

*Note*. $N = 40$.

**Non-parametric Statistics for Manipulation Check**

For each scenario, a chi-square goodness of fit test was run as a manipulation check. Distributions of the four possible responses from the participant perception question about the scenario are shown in Table 3.

**Uninformed surprise emergency for multi-engine aircraft**. A chi-square goodness of fit was run to test the null hypothesis that participant perception for uninformed surprise emergency for multi-engine aircraft occurred with equal probabilities. The test showed a significant discrepancy in the distributions of observed and expected frequencies, $\chi^2(3, N = 40) = 20.60, p < .001$. The scenario was surprising for 55% of the participants, while the second common response was surprising and startling (both), which accounted for 22.5% of the participants.

**Uninformed surprise and startle emergency for multi-engine aircraft**. A chi-square goodness of fit was run to test the null hypothesis that participant perception for uninformed surprise and startle emergency for multi-engine aircraft occurred with equal probabilities. The test showed a significant discrepancy in the distributions of observed and expected frequencies, $\chi^2(3, N = 40) = 15.40, p = .002$. The scenario was surprising and startling (both) for 50% of the participants, while the second common response was startling, which accounted for 22.5% of the participants.

**Informed emergency for multi-engine aircraft**. A chi-square goodness of fit was run to test the null hypothesis that participant perception for informed emergency for multi-engine aircraft occurred with equal probabilities. The test showed a significant discrepancy in the distributions of observed and expected frequencies, $\chi^2(3, N = 40) = 44.00, p < .001$. The scenario was neither surprising nor startling (neither) for 70% of the participants, while the second common response was surprising which accounted for 15% of the participants.

**Uninformed surprise emergency for single-engine aircraft**. A chi-square goodness of fit was run to test the null hypothesis that participant perception for uninformed surprise emergency for single-engine aircraft occurred with equal probabilities. The test showed a significant discrepancy in the distributions of observed and expected frequencies, $\chi^2(3, N = 40) = 35.40, p < .001$. The scenario was surprising for 65% of the participants, while the second common response was neither surprising nor startling (neither), which accounted for 17.5% of the participants.

**Uninformed surprise and startle emergency for single-engine aircraft**. A chi-square goodness of fit was run to test the null hypothesis that participant perception for uninformed surprise and startle emergency for multi-engine aircraft occurred with equal probabilities. The test showed a significant discrepancy in the distributions of observed and expected frequencies, $\chi^2(3, N = 40) = 16.80, p = .001$. The scenario was surprising and startling (both) for 50% of the participants, while the second common response was surprising which accounted for 25% of the participants.

**Informed emergency for single-engine aircraft**. A chi-square goodness of fit was run to test the null hypothesis that participant perception for informed emergency for single-engine aircraft occurred with equal probabilities. The test showed a significant discrepancy in the distributions of observed and expected frequencies, $\chi^2(3, N = 40) = 53.40, p < .001$. The scenario was neither surprising nor surprising (neither) for 75% of the participants, while the second common response was startling which accounted for 10% of the participants.

Since all six chi-square goodness of fit tests were significant, all the null hypotheses were rejected ($p < .05$); there were significant differences among the distributions of perception responses (surprise, startle, both, and neither) in different aircraft and emergency scenario conditions.

## Inferential Statistics

**Heart rate and respiration rate.** Separate 2 x 3 with-subjects ANOVAs were run to assess the effects of aircraft and emergency on HR and RR. The descriptive and inferential statistics are presented in Table 4 and Table 5.

Table 4

*Descriptive Statistics for HR and RR for Emergency Condition and Aircraft*

| | Multi-engine | | Single-engine | |
|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* |
| Heart Rate | | | | |
| Uninformed Surprise | 82.33 | 9.94 | 79.59 | 8.27 |
| Uninformed Surprise and Startle | 86.19 | 9.49 | 81.91 | 8.08 |
| Informed | 79.54 | 8.42 | 77.63 | 7.54 |
| Total | 82.69 | 8.78 | 79.71 | 7.43 |
| Respiration Rate | | | | |
| Uninformed Surprise | 22.36 | 3.09 | 20.90 | 2.72 |
| Uninformed Surprise and Startle | 23.72 | 3.18 | 21.44 | 2.83 |
| Informed | 20.80 | 2.84 | 19.90 | 2.63 |
| Total | 22.30 | 2.69 | 20.75 | 2.50 |

*Note*. $N = 40$. $M$ = mean; $SD$ = standard deviation.

Table 5

*Two-Way ANOVA Statistics for HR and RR*

| Dependent Variable | ANOVA | | | | | |
|---|---|---|---|---|---|---|
| | Effect | $MS_E$ | $F$ | $df$ | *p-value* | $\eta_p^2$ |
| Heart Rate | | | | | | |
| | A* | 532.82 | 9.03 | 1, 39 | .005 | 0.19[1] |
| | B** | 736.61 | 33.35[a] | 1.63, 63.61 | < .001 | 0.46[1] |
| | A x B* | 28.95 | 3.36 | 2, 78 | .040 | 0.08[2] |
| Respiration Rate | | | | | | |
| | A** | 143.38 | 22.95 | 1, 39 | < .001 | 0.37[1] |
| | B** | 137.57 | 34.64[a] | 1.45, 56.68 | < .001 | 0.47[1] |
| | A x B* | 9.72 | 5.16 | 2, 78 | .008 | 0.12[2] |

*Note*. $N$ = 40. A = aircraft; B = emergency; $MS_E$ = mean squared error; *df* = degrees of freedom; ANOVA = analysis of variance; $\eta_p^2$= partial eta square.
*$p < .05$. **$p < .01$. [a] Assumption of sphericity violated, thus an adjustment to *df* was made using Greenhouse Geisser. [1] large effect. [2] medium effect.

***Heart rate post hoc results for main effect of emergency.*** Using the Bonferroni post hoc, the mean heart rate of uninformed surprise emergency ($M$ = 80.96, $SD$ = 8.46) was significantly lower than the mean heart rate of uninformed surprise and startle emergency ($M$ = 84.05, $SD$ = 8.09, $p < .001$). The mean heart rate of informed emergency ($M$ = 78.59, $SD$ = 7.06) was significantly lower than the mean heart rate of uninformed surprise emergency ($p$ = .002) and the mean heart rate of uninformed surprise and startle emergency ($p < .001$). See Figure 3.

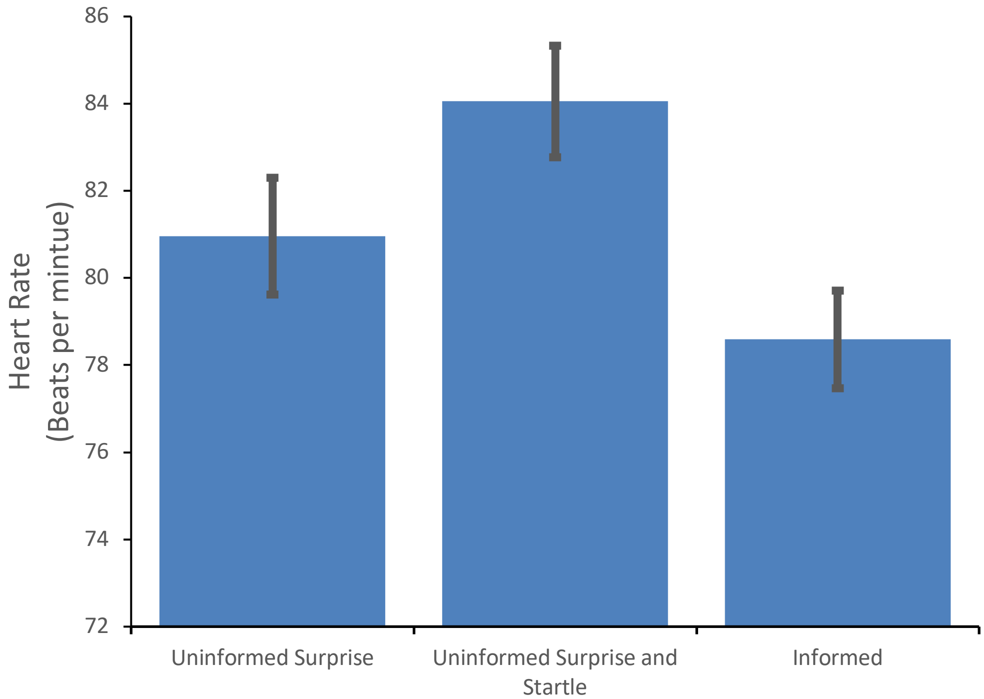


*Figure 3.* Mean heart rate with error bars (standard error of the mean) based on emergency.

***Heart rate post hoc results for interaction (aircraft*emergency).*** Using a paired sample t-test (testwise α = .017, Bonferroni adjustment), the mean heart rate of the uninformed surprise emergency for multi-engine aircraft was significantly higher than the mean heart rate of the uninformed surprise emergency for single-engine aircraft ($p$ = .016). Similarly, the mean heart rate of the uninformed surprise and startle emergency for multi-engine aircraft was significantly higher than the mean heart rate of the uninformed emergency for single-engine aircraft ($p < .001$). However, no significant difference in mean heart rate in the informed emergency condition was found between the single- and multi-engine aircraft ($p > .016$). See Figure 4.

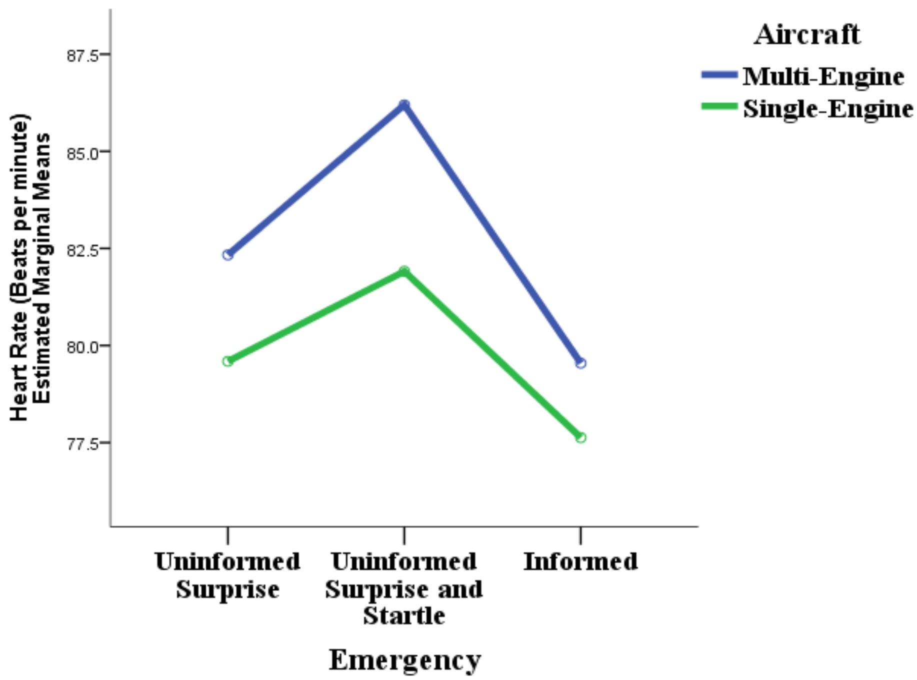


*Figure 4*. Mean heart rate for multi-engine and single-engine aircraft based on emergency.

***Respiration rate post hoc results for main effect of emergency.*** Using the Bonferroni post hoc, the mean respiration rate of uninformed surprise emergency ($M$ = 21.63, $SD$ = 2.62) was significantly lower than the mean respiration rate of uninformed surprise and startle emergency ($M$ = 22.58, $SD$ = 2.76, $p$ = .001). The mean respiration rate of informed emergency ($M$ = 20.40, $SD$ = 2.54) was significantly lower than the respiration rate of uninformed surprise emergency ($p$ < .001) and the mean respiration rate of uninformed surprise and startle emergency ($p$ < .001). No other significant differences were found ($p$ > .05). See Figure 5.

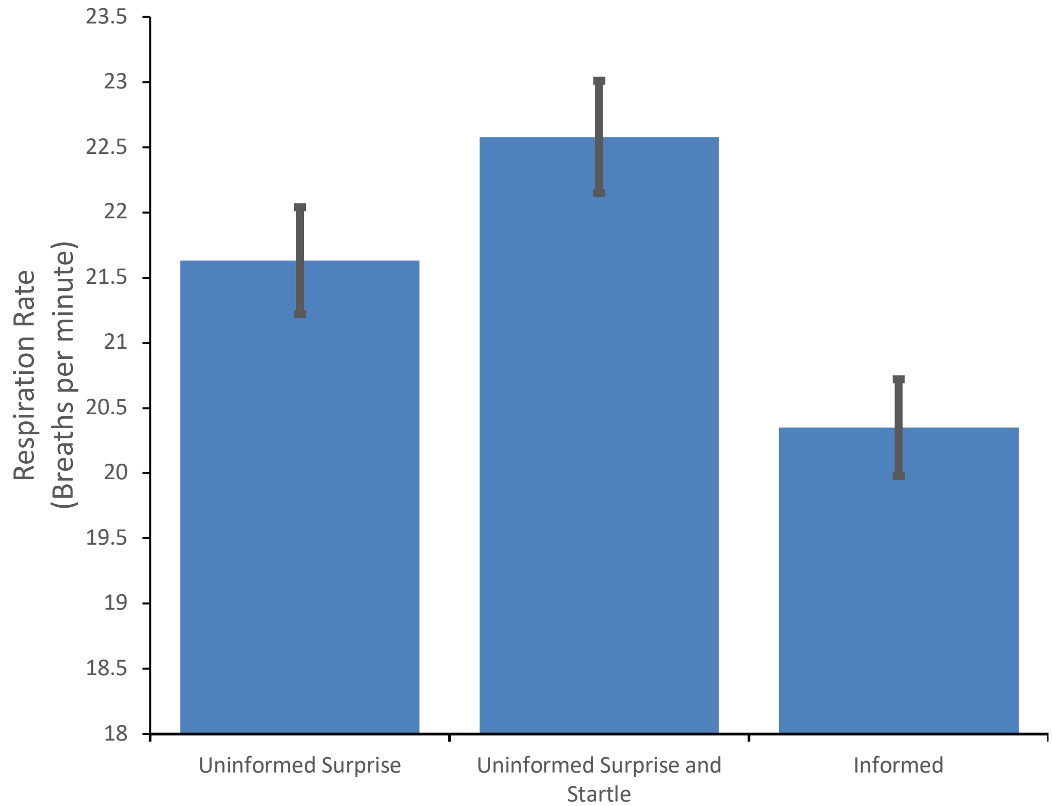


*Figure 5.* Mean respiration rate with error bars (standard error of the mean) based on emergency.

***Respiration rate post hoc results for interaction (aircraft*emergency).*** Using a paired sample *t*-test (testwise α = .016, Bonferroni adjustment), the mean respiration rate of the uninformed surprise emergency for multi-engine aircraft was significantly higher than the mean respiration rate of the uninformed surprise emergency for single-engine aircraft ($p = .001$). Similarly, the mean respiration rate of the uninformed surprise and startle emergency for multi-engine aircraft was significantly higher than the mean respiration rate of the uninformed emergency for single-engine aircraft ($p < .001$). However, no significant difference in mean respiration rate was found in the informed emergency condition between the single- and multi-engine aircraft ($p > .016$). These results are depicted in Figure 6.

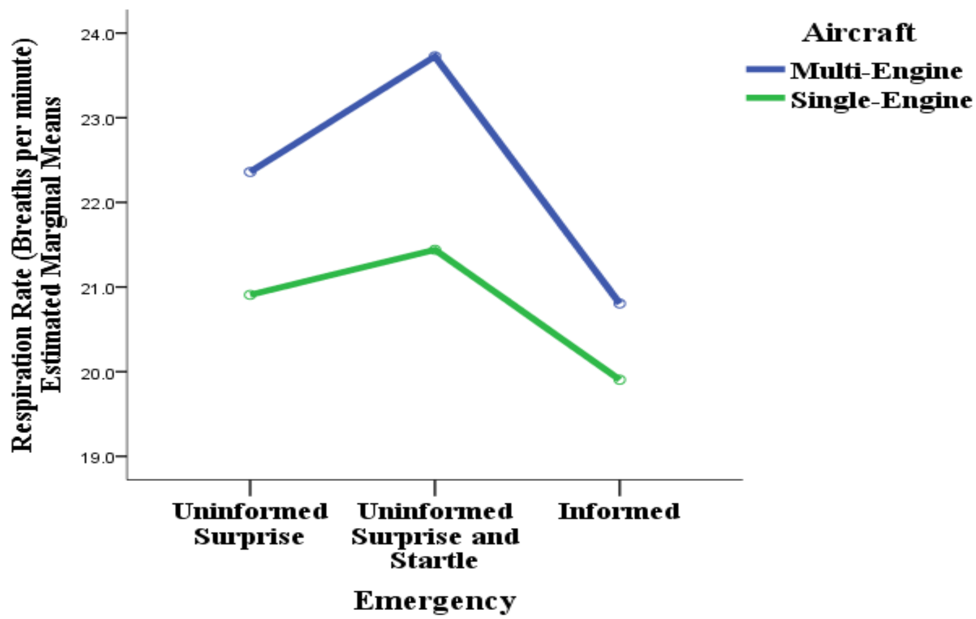


*Figure 6*. Mean respiration rate for multi-engine and single-engine aircraft based on emergency.

**Mental, physical, and temporal demand (subjective).** Three separate 2 x 3 repeated-measures ANOVAs were run to assess the effects of aircraft and emergency on the following dependent variables; (a) mental demand, (b) physical demand, and (c) temporal demand. The descriptive and inferential statistics are presented in Table 6 and 7 respectively.

Table 6

*Descriptive Statistics for Mental, Physical, and Temporal Demand for Emergency Condition and Aircraft*

| | Multi-engine | | Single-engine | |
|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* |
| Mental Demand | | | | |
| Uninformed Surprise | 13.27 | 4.69 | 11.08 | 4.50 |
| Uninformed Surprise and Startle | 14.72 | 4.50 | 13.85 | 4.57 |
| Informed | 11.15 | 4.63 | 8.30 | 4.64 |
| Total | 13.05 | 4.17 | 11.07 | 4.02 |
| Physical Demand | | | | |
| Uninformed Surprise | 9.33 | 4.42 | 8.05 | 4.81 |
| Uninformed Surprise and Startle | 12.53 | 4.96 | 9.80 | 4.65 |
| Informed Emergency | 10.05 | 5.05 | 7.05 | 4.72 |
| Total | 10.63 | 4.46 | 8.30 | 4.43 |
| Temporal Demand | | | | |
| Uninformed Surprise | 13.58 | 4.73 | 8.22 | 5.07 |
| Uninformed Surprise and Startle | 13.43 | 5.12 | 12.83 | 4.68 |
| Informed | 10.45 | 5.07 | 7.90 | 4.78 |
| Total | 12.48 | 4.27 | 9.65 | 4.04 |

*Note*. *N* = 40. *M* = mean; *SD* = standard deviation.

Table 7

*Two-Way ANOVA Statistics for Mental Demand, Physical Demand, and Temporal Demand*

| Dependent Variable | Effect | $MS_E$ | $F$ | $df$ | $p$-value | $\eta_p^2$ |
|---|---|---|---|---|---|---|
| Mental Demand | | | | | | |
| | A** | 234.04 | 12.58 | 1, 39 | .001 | 0.24[1] |
| | B** | 417.09 | 64.87 | 2, 78 | < .001 | 0.62[1] |
| | A x B* | 20.26 | 3.22 | 2, 78 | .045 | 0.08[2] |
| Physical Demand | | | | | | |
| | A** | 326.68 | 25.93 | 1, 39 | < .001 | 0.40[1] |
| | B** | 172.93 | 38.63 | 2, 78 | < .001 | 0.50[1] |
| | A x B* | 17.18 | 3.69 | 2, 78 | .029 | 0.09[2] |
| Temporal Demand | | | | | | |
| | A** | 481.67 | 17.31 | 1, 39 | < .001 | 0.31[1] |
| | B** | 313.72 | 37.79 | 2, 78 | < .001 | 0.49[1] |
| | A x B** | 114.02 | 9.32 | 2, 78 | < .001 | 0.19[1] |

*Note*. $N = 40$. A = aircraft; B = emergency; $MS_E$ = mean squared error; $df$ = degrees of freedom; ANOVA = analysis of variance; $\eta_p^2$= partial eta square.
*$p < .05$. **$p < .01$. [1] large effect. [2] medium effect.

***Mental demand post hoc results for main effect of emergency.*** Using the Bonferroni post hoc, the mean mental demand of uninformed surprise emergency ($M$ = 12.17, $SD$ = 4.06) was significantly lower than the mean mental demand of uninformed surprise and startle emergency ($M$ = 14.29, $SD$ = 4.09, $p < .001$). The mean mental demand of informed emergency ($M = 9.72$, $SD = 3.78$) was significantly lower than the mean mental demand of uninformed surprise emergency ($p < .001$) and the mean mental demand of uninformed surprise and startle emergency ($p < .001$). See Figure 7.

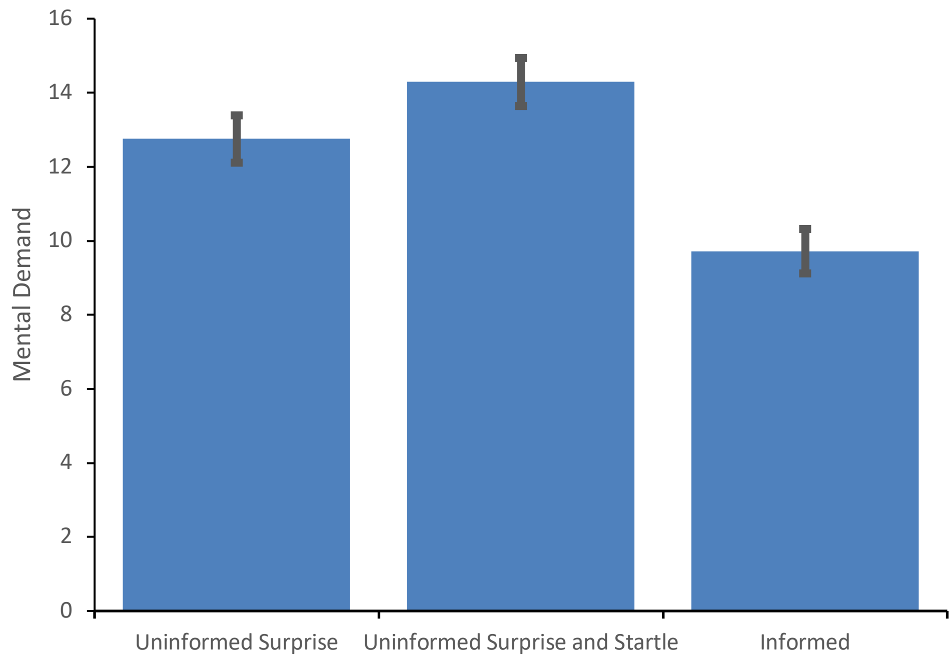


*Figure 7.* Mean mental demand with error bars (standard error of the mean) based on emergency.

***Mental demand post hoc results for interaction (aircraft*emergency).*** Using a paired sample *t*-test (testwise $\alpha = .016$, Bonferroni adjustment), the mean mental demand of the uninformed surprise emergency for multi-engine aircraft was significantly higher than the mean mental demand of the uninformed surprise emergency for single-engine aircraft ($p = .003$). Similarly, the mean mental demand of the informed emergency for multi-engine aircraft was significantly higher than the mean mental demand of the informed emergency for single-engine aircraft ($p = .002$). However, no significant difference in mean mental demand were found in the uninformed surprise and startle emergency condition between single- and multi-engine aircraft ($p > .016$). These results are depicted in Figure 8.

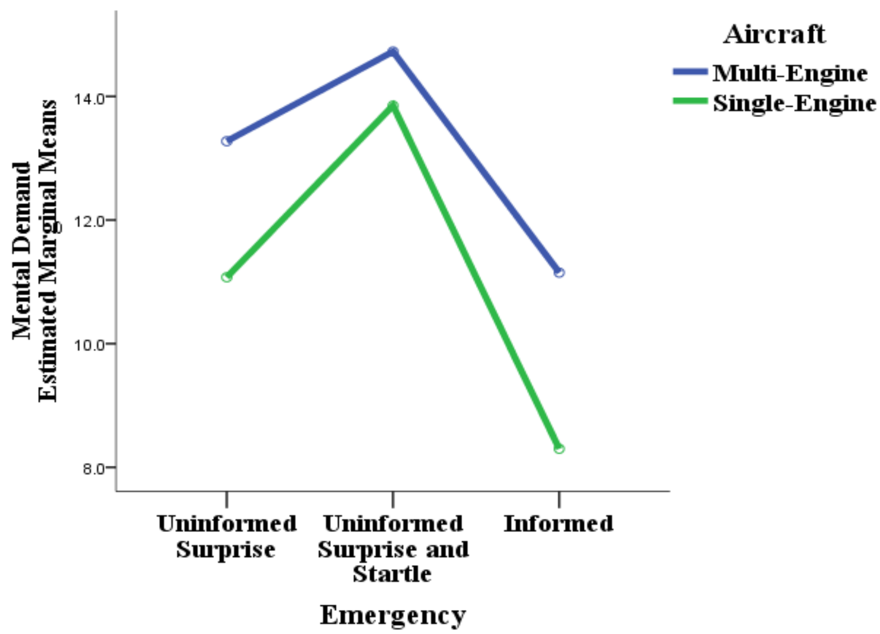


*Figure 8*. Mean mental demand for multi-engine and single-engine aircraft based on emergency.

***Physical demand post hoc results for main effect of emergency.*** Using the Bonferroni post hoc, the mean physical demand of uninformed surprise emergency ($M = 8.69$, $SD = 4.18$) was significantly lower than the mean physical demand of uninformed surprise and startle emergency ($M = 11.16$, $SD = 4.37$, $p < .001$). The mean physical demand of informed emergency ($M = 8.55$, $SD = 4.56$) was significantly lower than the mean physical demand of the uninformed surprise and startle emergency ($p < .001$). There was no significant difference in mean physical demand between the uninformed surprise emergency and the informed emergency ($p > .05$). See Figure 9.

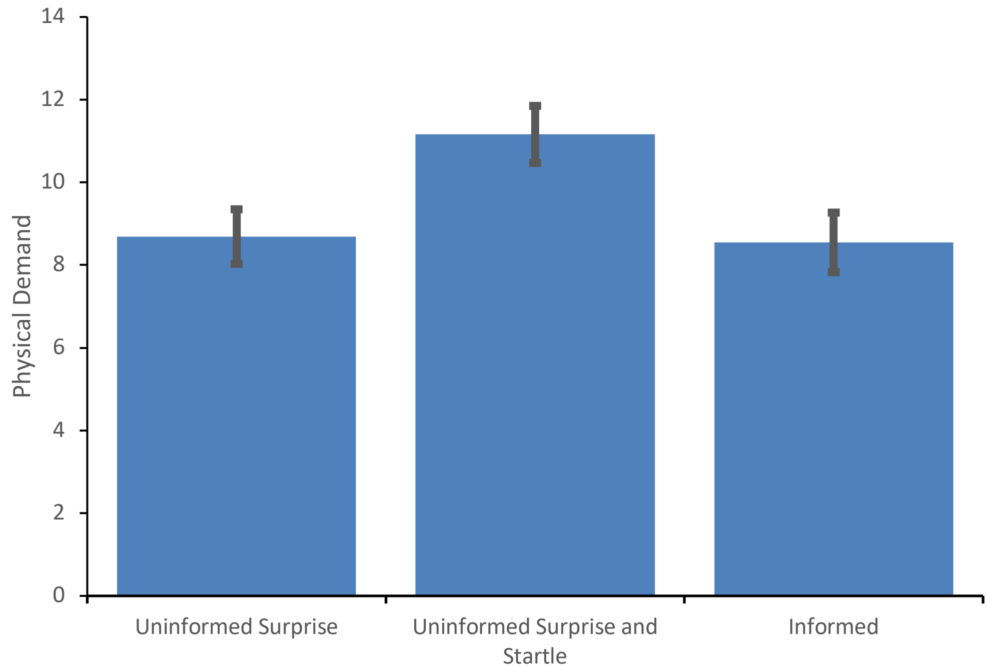


*Figure 9.* Mean physical demand with error bars (standard error of the mean) based on emergency.

***Physical demand post hoc results for interaction (aircraft*emergency).*** Using a paired sample *t*-test (testwise $\alpha = .016$, Bonferroni adjustment), the mean physical demand of the uninformed surprise and startle for multi-engine aircraft was significantly higher than the mean physical demand of the uninformed surprise and startle and for single-engine aircraft ($p < .001$). Similarly, mean physical demand of the informed emergency for multi-engine aircraft was significantly higher than the mean physical demand of the informed emergency for single-engine aircraft ($p < .001$). However, there was no significant difference between the aircraft for the uninformed surprise condition ($p > .016$). These results are depicted in Figure 10.

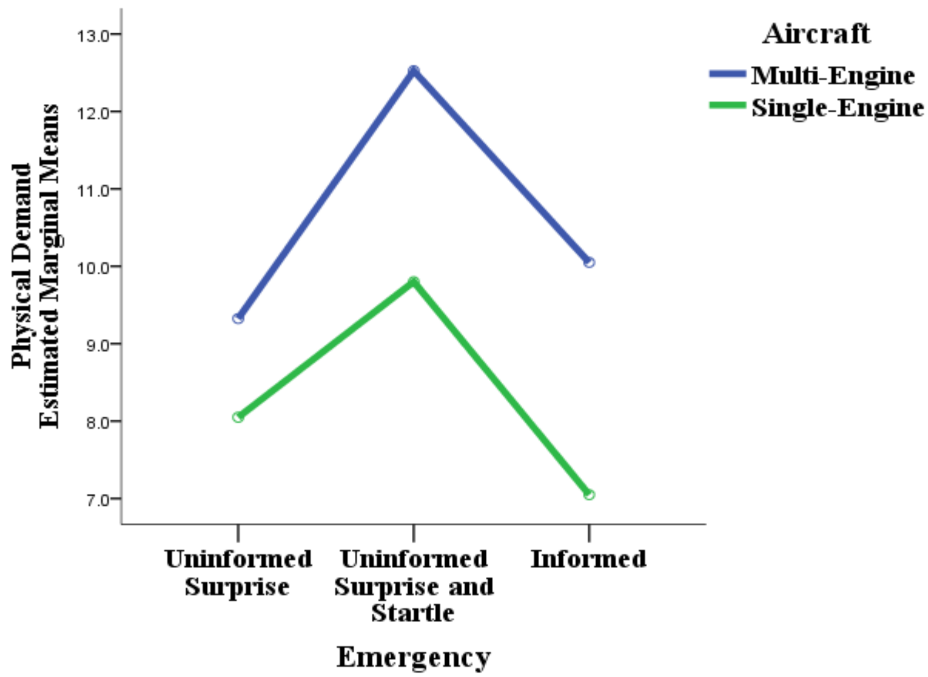


*Figure 10*. Mean physical demand for multi-engine and single-engine aircraft based on emergency.

***Temporal demand post hoc results for main effect of emergency.*** Using the Bonferroni post hoc, the mean temporal demand of uninformed surprise emergency ($M$ = 10.90, $SD$ = 3.89) was significantly lower than the mean temporal demand of uninformed surprise and startle emergency ($M$ = 13.12, $SD$ = 3.63, $p < .001$). The mean temporal demand of informed emergency ($M$ = 9.17, $SD$ = 4.23) was significantly lower than the mean temporal demand of uninformed surprise emergency ($p = .002$) and the mean temporal demand of uninformed surprise and startle emergency ($p < .001$). See Figure 11.

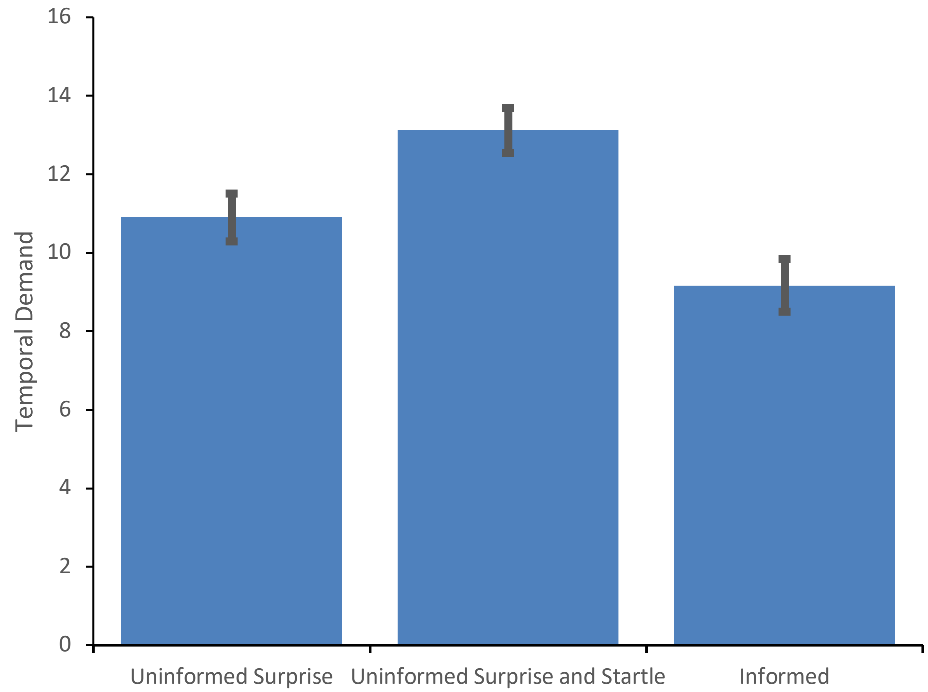


*Figure 11.* Mean temporal demand with error bars (standard error of the mean) based on emergency.

***Temporal demand post hoc results for interaction (aircraft*emergency*).** Using a paired sample *t*-test (testwise $\alpha = .016$, Bonferroni adjustment), the mean temporal demand of the uninformed surprise emergency for multi-engine aircraft was significantly higher than the mean temporal demand of the uninformed surprise emergency for single-engine aircraft ($p < .001$). Similarly, the mean temporal demand of the informed emergency for multi-engine aircraft was significantly higher than the mean temporal demand of the informed emergency for single-engine aircraft ($p = .003$). However, no significant differences were found in mean temporal demand in the uninformed surprise and startle emergency condition between the single- and multi-engine aircraft ($p > .016$). These results are depicted in Figure 12.

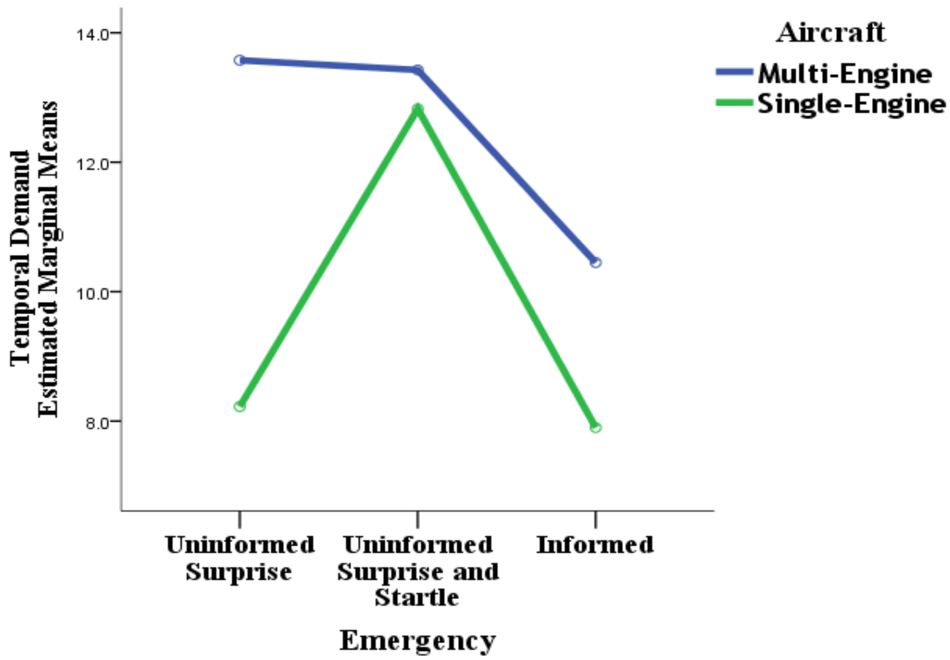


*Figure 12*. Mean temporal demand for multi-engine and single-engine aircraft based on emergency.

**Subjective performance, effort, and frustration.** Three separate 2 x 3 repeated-measures ANOVAs were run to assess the effects of aircraft and emergency on the following dependent variables; (a) subjective performance, (b) effort, and (c) frustration. The descriptive and inferential statistics are presented in Table 8 and 9 respectively.

Table 8

*Descriptive Statistics for Subjective Performance, Effort, and Frustration for Emergency Condition and Aircraft*

| | Multi-engine | | Single-engine | |
|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* |
| Subjective Performance | | | | |
| Uninformed Surprise | 15.20 | 4.78 | 13.03 | 4.91 |
| Uninformed Surprise and Startle | 16.70 | 3.90 | 14.33 | 3.85 |
| Informed | 11.98 | 5.08 | 6.67 | 4.45 |
| Total | 14.63 | 3.19 | 11.34 | 3.29 |
| Effort | | | | |
| Uninformed Surprise | 12.70 | 4.40 | 10.67 | 4.42 |
| Uninformed Surprise and Startle | 15.28 | 4.37 | 12.67 | 4.53 |
| Informed Emergency | 11.67 | 3.93 | 9.15 | 4.43 |
| Total | 13.22 | 3.35 | 10.83 | 3.88 |
| Frustration | | | | |
| Uninformed Surprise | 13.15 | 5.33 | 9.75 | 5.67 |
| Uninformed Surprise and Startle | 14.78 | 5.45 | 12.18 | 5.50 |
| Informed | 70.23 | 5.28 | 7.03 | 4.65 |
| Total | 12.72 | 4.70 | 9.65 | 4.51 |

*Note*. *N* = 40. *M* = mean; *SD* = standard deviation.

Table 9

*Two-Way ANOVA Statistics for Subjective Performance, Effort, and Frustration*

| Dependent Variable | ANOVA | | | | | |
|---|---|---|---|---|---|---|
| | Effect | $MS_E$ | $F$ | $df$ | $p$-value | $\eta_p^2$ |
| Subjective Performance | | | | | | |
| | $A^{**}$ | 646.82 | 23.01 | 1, 39 | < .001 | 0.37[1] |
| | $B^{**}$ | 842.20 | 66.00 | 2, 78 | < .001 | 0.63[1] |
| | $A \times B^{*}$ | 61.20 | 3.60 | 2, 78 | .032 | 0.08[2] |
| Effort | | | | | | |
| | $A^{**}$ | 340.82 | 18.83 | 1, 39 | < .001 | 0.33[1] |
| | $B^{**}$ | 260.66 | 27.88 | 2, 78 | < .001 | 0.41[1] |
| | $A \times B^{ns\ \&\ a}$ | 2.26 | 0.25 | 1.73, 67.35 | > .05 | - |
| Frustration | | | | | | |
| | $A^{**}$ | 564.27 | 39.93 | 1, 39 | < .001 | 0.51[1] |
| | $B^{**}$ | 474.72 | 44.16 | 2, 78 | < .001 | 0.53[1] |
| | $A \times B^{ns}$ | 3.47 | 0.32 | 2, 78 | > .05 | - |

*Note*. $N = 40$. A = aircraft; B = emergency; $MS_E$ = mean squared error; *df* = degrees of freedom; ANOVA = analysis of variance; $\eta_p^2$= partial eta square.
[a] Assumption of sphericity violated, thus an adjustment to *df* was made using Greenhouse Geisser. [ns] Not significant ($p > .05$). [1] large effect. [2] medium effect.

***Subjective performance post hoc results for main effect of emergency.*** Using the Bonferroni post hoc, the mean subjective performance of uninformed surprise emergency ($M = 14.11$, $SD = 3.19$) was significantly lower than the mean performance of uninformed surprise and startle emergency ($M = 15.51$, $SD = 2.60$, $p$ .029). The mean subjective performance of informed emergency ($M = 9.32$, $SD = 3.64$) was significantly lower than the mean subjective performance of uninformed surprise emergency ($p < .001$) and the mean subjective performance of uninformed surprise and startle emergency ($p < .001$). See Figure 13.

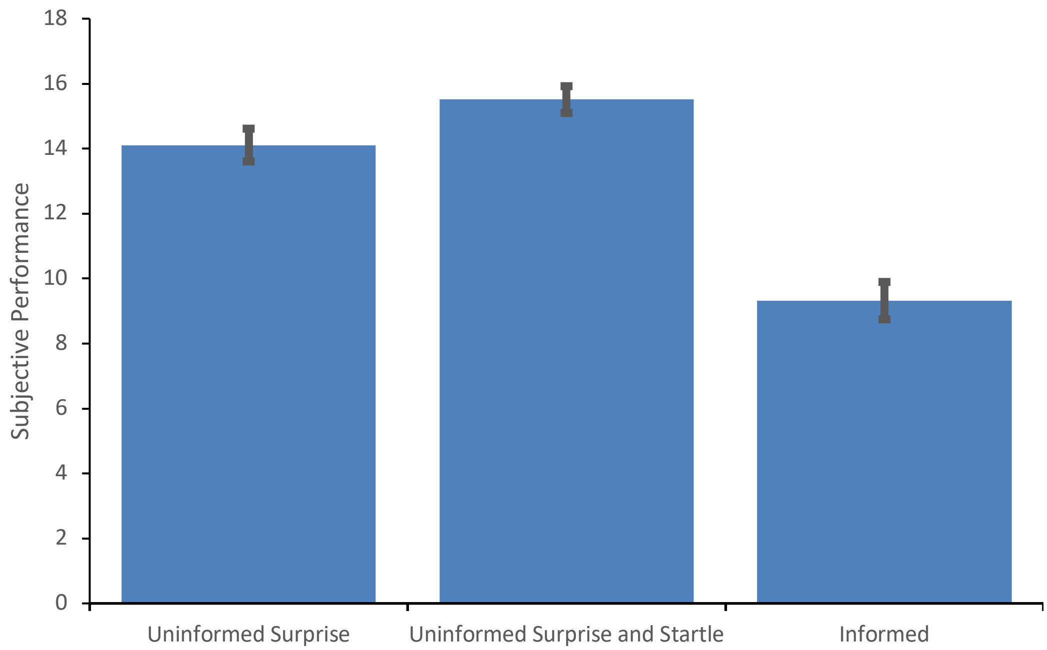


*Figure 13.* Mean subjective performance with error bars (standard error of the mean) based on emergency.

***Subjective performance post hoc results for interaction (aircraft*emergency*).** Using a paired sample *t*-test (testwise α = .016, Bonferroni adjustment), the mean subjective performance of the uninformed surprise and startle emergency for multi-engine aircraft was significantly higher than the mean subjective performance of the uninformed surprise and startle emergency for single-engine aircraft ($p = .013$). Similarly, the mean subjective performance of the informed emergency for multi-engine aircraft was significantly higher than the mean subjective performance of the informed emergency for single-engine aircraft ($p < .001$). However, there was no significant difference in mean subjective performance in the uninformed surprise emergency condition between the single- and multi-engine aircraft ($p > .016$). These results are depicted in Figure 14.

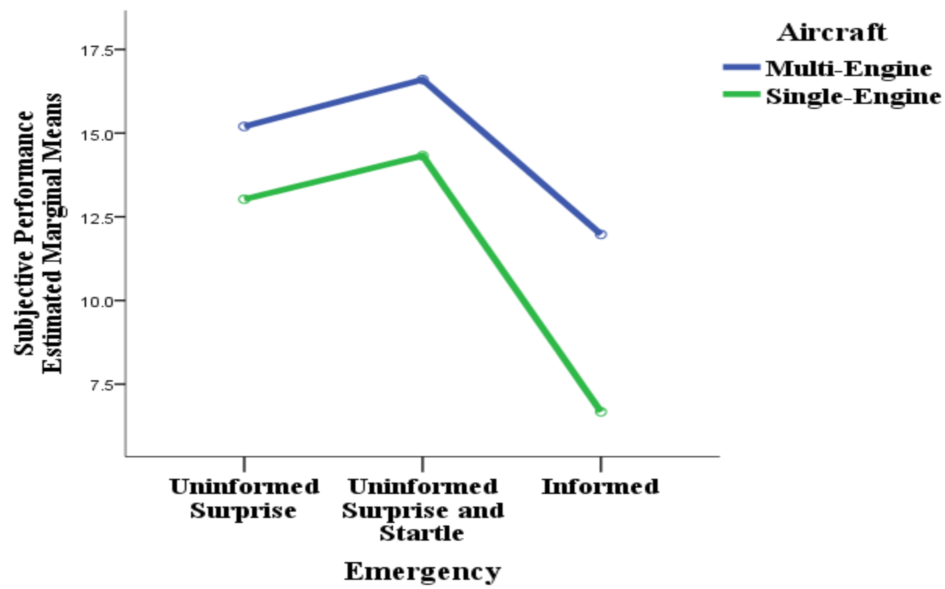


*Figure 14*. Mean subjective performance for multi-engine and single-engine aircraft based on emergency.

***Effort post hoc results for main effect of emergency.*** Using the Bonferroni post hoc, the mean effort of uninformed surprise emergency ($M = 11.69$, $SD = 3.59$) was significantly lower than the mean effort of uninformed surprise and startle emergency ($M = 13.97$, $SD = 3.61$, $p < .001$). The mean effort of informed emergency ($M = 10.41$, $SD = 3.70$) was significantly lower than the mean effort of uninformed surprise emergency ($p = .033$) and the mean effort of uninformed surprise and startle emergency ($p < .001$). See Figure 15.

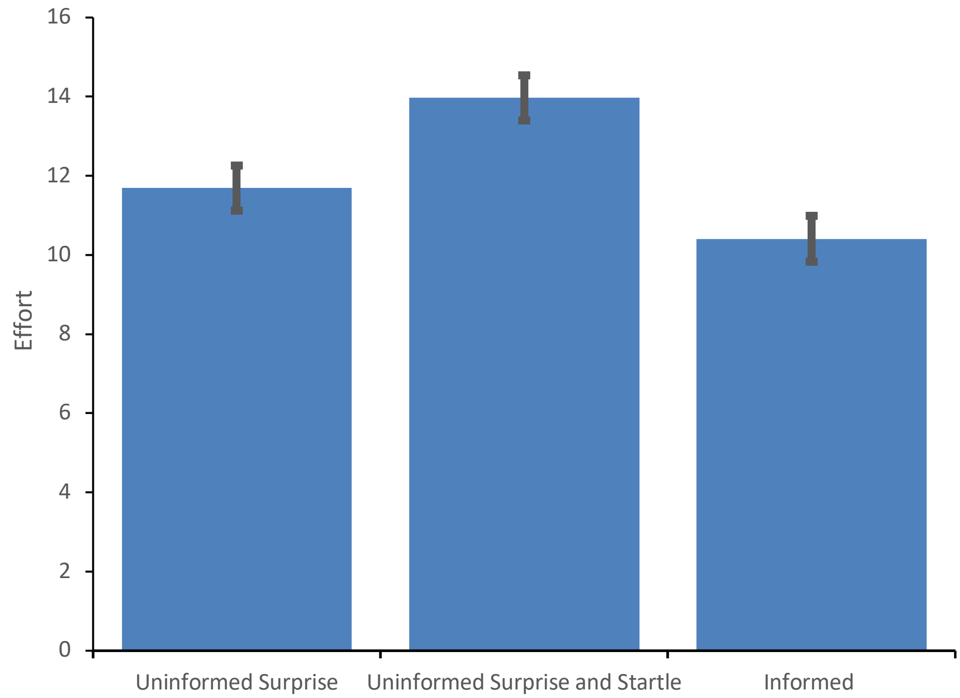


*Figure 15.* Mean effort with error bars (standard error of the mean) based on emergency.

***Frustration post hoc results for main effect of emergency.*** Using the Bonferroni post hoc, the mean frustration of uninformed surprise emergency ($M$ = 11.45, $SD$ = 4.76) was significantly lower than the mean frustration of uninformed surprise and startle emergency ($M$ = 13.47, $SD$ = 4.80, $p$ = .001). The mean frustration of the informed emergency ($M$ = 8.62, $SD$ = 4.69) was significantly lower than the mean frustration of the uninformed surprise emergency ($p < .001$) and the mean frustration of uninformed surprise and startle emergency ($p < .001$). See Figure 16.

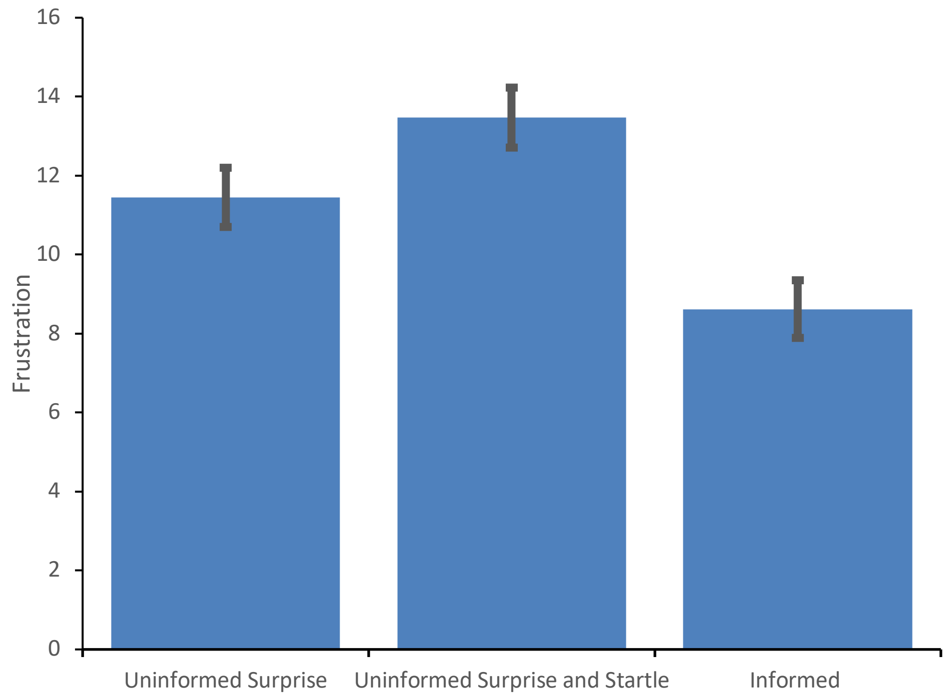


*Figure 16.* Mean frustration with error bars (standard error of the mean) based on emergency.

**Checklist Compliance Results for Single-Engine Aircraft**

A one-way repeated measures ANOVA was conducted to test the null hypothesis that there would be no significant difference in the number of checklist steps followed among emergency conditions in the single-engine aircraft condition. The assumption of sphericity was tested. Mauchly's test of sphericity was not significant ($p > .05$). The number of engine failure checklist steps followed between the emergencies significantly varied, $F(2, 78) = 106.10$, $p < .001$, $\eta^2 = .73$ (large effect size). The Bonferroni post hoc showed that the mean number of checklist steps followed for uninformed surprise emergency was significantly greater than the mean number of checklist steps followed for uninformed surprise and startle emergency ($p < .001$). The mean number of checklist steps followed for an informed emergency was significantly greater than the mean number of checklist steps followed for uninformed surprise emergency ($p < .001$) and uninformed surprise and startle emergency ($p < .001$). The means and standard deviations of the checklist steps followed based on emergency type are shown in Table 7.

Table 7

*Descriptive Statistics for Mean Number of Checklist Steps Followed for Single-Engine Aircraft*

| Emergency Condition | *M* | *SD* |
|---|---|---|
| Uninformed Surprise Emergency | 3.40 | 2.16 |
| Uninformed Surprise and Startle Emergency | 1.82 | 1.63 |
| Informed Emergency | 6.80 | 1.80 |

*Note*. $N = 40$. $M$ = Mean; $SD$ = Standard Deviation.

**Flight Performance Results for Multi-Engine Aircraft**

A one-way repeated-measures ANOVA was conducted to test the null hypothesis that there would be no significant difference in altitude deviation among the emergency conditions. The assumption of sphericity was tested. Mauchly's test of sphericity was not significant ($p > .05$). The altitude deviation between the emergencies significantly varied, $F(2, 78) = 67.34$, $p < .001$, $\eta^2 = .63$ (large effect size). The Bonferroni post hoc indicted that the mean altitude deviation for uninformed surprise emergency was significantly less than the mean altitude deviation for uninformed surprise and startle emergency ($p = .043$). The mean altitude deviation for an informed emergency was significantly less than the mean altitude deviation for uninformed surprise emergency

($p < .001$) and uninformed surprise and startle emergency ($p < .001$). The means and standard deviations of the altitude deviation for emergencies are depicted in Table 8.

Table 8

*Descriptive Statistics for Altitude Deviation for Multi-Engine Aircraft*

| Emergency Condition | *M* | *SD* |
|---|---|---|
| Uninformed Surprise Emergency | 122.12 | 49.93 |
| Uninformed Surprise and Startle Emergency | 147.77 | 46.43 |
| Informed Emergency | 27.35 | 50.32 |

*Note*. $N = 40$. $M$ = Mean; $SD$ = Standard Deviation.

# Chapter V

## Discussion, Conclusions, and Recommendations

The results are discussed in this chapter by giving a wider insight into the possible reasons for the findings. Since the topic under investigation is a relatively new area of research, the researcher also suggests recommendations for future studies. This chapter includes personal communication with the participants and experienced researchers to help with the interpretation of the results.

### Discussion

**Manipulation check.** A total of 22 participants (55 %) found the uninformed surprise condition for the single-engine aircraft surprising, while 28 participants (65 %) found it surprising for the multi-engine aircraft. However, 20 participants (50 %) found the uninformed surprise and startle condition to be both surprising and startling for the single-engine and the multi-engine aircraft. As discussed in Chapter 2, surprise and startle are often used interchangeably, and further, the term *surprise* is sometimes used interchangeably with *startle*. For example, Rivera et al. (2014) in a review of the ASRS database found that the term *startle* is often not used to refer to startle but surprise. Findings from Rivera et al. (2014) are consistent with this study where in the uninformed surprise and startle condition, 8 participants (20 %) experienced surprise for the multi-engine aircraft, while 10 participants (25 %) experienced surprise for the single-engine aircraft. Surprise and startle are different constructs with different causes and effects. Hence it is very important to distinguish between surprise and startle, otherwise, it can lead to only partial understanding of the effects of surprise and startle on pilots during unexpected events. If these effects are not distinguished and understood, it can potentially void the benefits of any training scenarios developed based on partial understanding.

Twenty-eight participants (70 %) found the informed condition neither surprising nor startling for the multi-engine aircraft, while 30 participants (75%) found the informed condition neither surprising nor startling for the single-engine aircraft. For both aircraft, the results were evident to show that the manipulation had worked on most participants. However, a potential limitation of these results was that they are based on participant perception and cannot reflect on how others in the population would perceive these conditions.

**Vital signs: Heart rate and respiration rate**. As expected and consistent with Bruna et al. (2018), respiration rate (RR) was higher for an uninformed emergency compared to an informed emergency. Similarly, heart rate (HR) was also higher for the uninformed emergency compared to an informed emergency, however, the results were not consistent with Landman et al. (2017b), where no significant differences were found between an informed and uninformed surprise emergency. The researcher would remind the readers that Bruna et al. (2018) and Landman et al. (2017b) had airline pilots as their sample; however, the current study had commercial pilots as its sample. Further, the current study was the first to explore the effects of startle and surprise on commercial pilots, and the results can be potentially generalized to the population. The results categorized the uninformed emergency conditions as an uninformed surprise and uninformed startle and surprise; and found that the HR and RR are higher for the uninformed startle and surprise condition. The loud bang and thunder noise were startling for most participants, especially participants flying the multi-engine aircraft where the average HR was 86.06 beats per min compared to a HR of 81.91 for the single-engine aircraft.

The mean HR of pilots flying the multi-engine aircraft was higher than the single-engine aircraft; see Table 4. Most participants agreed that the multi-engine aircraft was harder to fly as they had to look at multiple instruments while flying ILS conditions. An increased heart or respiration rate does not necessarily mean that the pilots were startled or surprised, which is why the researcher asked all participants their perception of the task. However, the manipulation check showed that pilots' subjective perceptions were generally consistent with the startle and surprise that the conditions were designed to induce for each aircraft. One participant felt his "heart beating faster" when he heard the loud bang, while another participant had sudden body movements when he heard the thunder sound. The sudden body movements when the sounds were played were consistent across most participants, something the researcher visually observed. The results were also consistent with how the FAA and the EASA defined startle and surprise, where startle is something associated with higher heart rate and surprise being something different than the expectation.

**NASA-TLX**. The multi-engine aircraft was mentally and physically demanding for most pilots. The researcher, based on his interaction with participants, offers one possible explanation for this effect: since most commercial pilots fly the Cessna (single-engine) aircraft regularly, they are more comfortable with flying a Cessna as they have more hours on that aircraft. It is possibly due to this reason that the mental and physical workload for an informed emergency was significantly less for the single-engine aircraft.

The mental demand was not different between the aircraft for the surprise and startle condition. The loud bang and thunder noise were meant to startle the pilots, and startle is associated with fear. It is highly unlikely that the fear due to the startling event will be different if the pilot was flying a different aircraft.

The multi-engine aircraft was more physically demanding as the pilots had to initiate a go-around; also in the informed emergency condition, the physical demand scores were higher than in the uninformed surprise condition. Initiating a go-around with engine failure is a physically demanding task, even if the condition is informed. The temporal demand for the multi-engine aircraft was higher for the surprise emergency than the surprise and startle emergency, which was something not expected. For the multi-engine aircraft, 12 participants crashed the aircraft in the surprise condition, while 22 participants crashed the aircraft in the surprise and startle condition. Based on examining the data again, it was found that 11 participants that crashed the multi-engine aircraft in the surprise condition also crashed the aircraft again in the surprise and startle condition. These participants did not perceive the temporal demand to be high for the surprise and startle condition based on mere exposure as they were experiencing a similar outcome the second time around, so they had acclimated to the situation.

Similarly, like mental demand, there was no difference in the temporal demand between the aircraft for the uninformed surprise and startle emergency. It can be argued that the initial behavioral response to fear would not vary with respect to whether the pilot is flying a single- or multi-engine aircraft. For example, if a person hears a gunshot which made him jump from his chair thinking what happened (mental demand) and run (temporal demand), it does not really matter if the person was at home or at work.

Subjective performance (self-assessed), as measured using the NASA-TLX, had the highest score among the six subscales for both the aircraft. On average, most participants rated their subjective performance as not up to their own standards except for the informed condition for the single-engine aircraft, which had the lowest score among all subscales for the informed condition. One participant, after flying the surprise and startle condition for the multi-engine aircraft said, “I know how to feather and initiate missed approach, but the loud bang interrupted me.” As per another participant, “the thunder and lightning forced me to land quick and I was not looking at my altitude”, when asked why, the participant replied, “I guess I was sure that I will see the runway, but I never did.” The author believes that most pilots were very critical of their performance; some even apologized and offered to re-fly the scenario. Based on personal communication with the participants, the study found that most pilots that flew the single-engine informed condition found it relatively straight forward, as it is an emergency that they have flown multiple times in a simulator. More specifically, participants that flew this scenario as their last flight were happy that they ended the experiment on a good note; one participant said, “finally I was able to do what I wanted to do, and there was no disturbance.”

Effort and frustration significantly varied across the aircraft and emergency conditions. For frustration, the researcher had not expected significant interactions for frustration, which was corroborated by discussions with participants. One participant suggested that “for me flying one aircraft over another is not frustrating as I need to build my hours.” However, the researcher expected to find a significant interaction for effort, especially after a significant interaction was found for mental and physical demand.

There is no obvious explanation for this result, as some participants verbally suggested that the multi-engine aircraft took more effort than the single-engine aircraft. It is one of those cases where the difference was just not statistically significant, though the mean effort for the multi-engine aircraft for all three emergency conditions was higher than the single-engine aircraft; see Table 7.

**Flight performance for single-engine aircraft (checklist steps followed)**. Since the participants were not expecting an engine failure in the surprise condition, they may not have been mentally prepared for the checklist, which was evident from the higher mean number of checklist steps followed in the informed condition. The surprise and startle condition further decreased the performance; 26 participants (65 %) in that condition forgot to enrich the mixture before restarting the engine, while 20 participants (50 %) failed to identify the landing field. The performance measured using the checklist was consistent with the self-assessed performance for both the aircraft. Engine failure training should incorporate startle and surprise as a factor. The results do signify that having an informed engine-failure will result in high flight performance, but present training will not necessarily transfer to situations where the pilots are faced with the same emergency unexpectedly.

**Flight performance for multi-engine aircraft (altitude deviation)**. Similar to the single-engine aircraft, altitude deviation performance was better in the informed condition, followed by the surprise condition, and the surprise and startle condition. More than half the participants ($N = 22$) crashed their aircraft in the surprise and startle condition; the engine-failure and the startle manipulation likely degraded conditions too much for the participants. However, one participant who was not happy with his performance said, “I was trying to find the runway and did not monitor my instruments for about 10 to 15 s”, in these 10 to 15 s the aircraft had an engine failure, and the altitude suddenly dropped, which led to the crash. All this went unnoticed by the participant, who was busy trying to see the runway visually.

**Practical Implications**

The effects of startle and surprise are well documented; however, this topic is under researched in the aviation industry. A study done on commercial pilots found that simulator training can be helpful in mitigating the effects of startle (Gillen, 2016). Similarly, studies also found that vital signs (respiration rate) and skin conductance increase in surprise conditions (Bruna et al., 2018; Landman et al., 2017b). Based on these studies, the research fraternity can agree that startle and surprise can affect performance and vital signs, and that training can possibly help mitigate the negative effects, which include inappropriate response during unexpected events not consistent with training. However, the workload was not looked at by any of the previous studies. Similarly, which simulator scenarios would be used for training, and for what aircraft? This study tried answering those questions for commercial pilots, which represent 21.8 % of the total pilot population in the U.S. It was found that flying a multi-engine aircraft during surprise and startle events can lead to higher workload, HR, and RR. Similarly, flying a single-engine aircraft in the uninformed surprise and startle condition can lead to higher HR, RR, and workload.

The simulator scenarios proposed in this study can be potentially be used for startle and surprise training for commercial pilots. The researcher believes that classroom training along with simulator training can help to mitigate the adverse effects of unexpected events on pilot performance. The key factor in the successful implementation of these simulator scenarios is that the pilots be uninformed about the emergency. If pilots fly an informed emergency, their performance will be better (Bhana, 2010). The results of this study substantiated that claim, where pilots' flight performance was better in informed emergency conditions and more than 70% of the participants were neither surprised nor startled while flying the informed emergency condition. However, the flight performance deteriorated in the uninformed surprise condition and worsened in the uninformed surprise and startle condition.

A potential contribution of this study was that the results suggested that heart rate and respiration rate can be used as a physiological measurement for startle and surprise. Further studies can record and evaluate heart rate and respiration rate to ascertain if the simulator scenarios were startling or surprising as intended. Also, based on the results, the researcher suggests that startle and surprise are not interchangeable terms in the aviation industry, and that the research fraternity considers this concept. It is important to understand that contradicting research on this topic cannot help with paving the way for potential federal regulations.

This study identified that aircraft could be a factor during startle and surprise events. Further, flying a surprise and startle emergency is more challenging than a surprise emergency. Keeping the results of this study in mind, more simulator scenarios should be proposed, something the FAA also suggested in its AC (2017).

**Conclusions**

For commercial pilots, the type of aircraft they are flying can impact their performance, vital signs (HR and RR), and workload during surprising and startling events. It was evident from the results that HR and RR varied in response to emergency conditions in similar patterns and increased from uninformed surprise conditions to the uninformed surprise and startle condition. The workload did increase significantly and was dependent on the aircraft and type of emergency. It was mentally and physically harder to fly a multi-engine aircraft, as evidenced by higher levels of frustration and effort. The key result of this study highlighted that having pilots fly informed emergency scenarios is not a good idea because it might make the training predictable; something with which Bhana (2010) agreed. The results of this study found that flight performance was better when the pilots were flying informed emergency scenarios compared to when they were flying uninformed emergency scenarios. The scenarios proposed in this study were surprising and startling for most pilots and can be used for training pilots, provided they are uninformed.

## Recommendations

**Vital signs.** This study, along with past studies, recorded and evaluated vital signs, which included HR, RR, and blood pressure. Past studies also recorded skin conductance and heart rate variability. Ideally, future studies should focus on the same measures for vital signs (HR, RR, blood pressure, and temperature) and try to validate the designs of previous studies with a different population. However, the researcher highly suggests that future studies should focus on electroencephalography evaluating the alpha, beta, and gamma waves. Electroencephalography is used in physiological research to evaluate the processing of complex stimuli (Biasiucci, Franceschiello, & Murray, 2019). The results of these futures studies can help identify the brain wave patterns during startling and surprising events on the flight deck, something no previous study has evaluated. Finally, electromyography is also a validated measure for startle (Blumenthal et al., 2005; Khemka, Tzovara, Gerster, Quednow, & Bach, 2017), but has not been used in any published aviation-related study so far.

**Workload using the NASA-TLX.** The current study was the first to assess pilot workload using NASA-TLX for unexpected events that can surprise and startle pilots. Future studies on this topic should use NASA-TLX to assess the pilot perception of the workload. The researcher believes that most studies conducted on startle and surprise from an aviation perspective did not employ any mechanisms to gather pilot perception of the task. Future studies can ask pilots their perception of the tasks as a manipulation check to establish if the scenario were surprising, startling, both or neither.

**Sample size.** Since few results were over-powered, the researcher recommends that future studies should conduct power analysis, preferably Beta testing, to estimate effect size for estimating sample size. The results for the six NASA-TLX factors, heart rate, respiration rate were over-powered for the main effect of aircraft and emergency scenarios. Similarly, the results for flight performance for both aircraft were also over-powered.

The results for all interactions had adequate power, so it is important to understand that if the goal is to evaluate the interaction between aircraft and emergency, then a sample of 40 is appropriate. However, if future studies only want to evaluate aircraft or emergency, then a sample of 40 would result in over-powered results.

## References


Bhana, H. (2010). Correlating boredom proneness and automation complacency in modern airline pilots. *Collegiate Aviation Review*, *28*(1), 9–24. Retrieved from http://search.proquest.com.ezproxy.libproxy.db.erau.edu/docview/859874566?accountid=27203

Biasiucci, A., Franceschiello, B., & Murray, M. M. (2019). Electroencephalography. *Current Biology*, *29*(3), R80-R85. doi:10.1016/j.cub.2018.11.052

Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S. A., Lipp, O. V., & van Boxtel, A. (2005). Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology*, *42*(1), 1-15. doi:10.1111/j.1469-8986.2005.00271.x

Boeing. (2019). Statistical summary of commercial jet airplane accidents. Retrieved from https://www.boeing.com/resources/boeingdotcom/company/about_bca/pdf/statsum.pdf

Boubin, J. G., Rusnock, C. F., & Bindewald, J. M. (2017). Quantifying compliance and reliance trust behaviors to influence trust in human-automation teams. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*, 750–754. https://doi.org/10.1177/1541931213601672

Bruna, O., Levora, T., & Holub, J. (2018). Assessment of ECG and respiration recordings from simulated emergency landings of ultra light aircraft. *Scientific Reports*, *8*(1), 7232-10. doi:10.1038/s41598-018-25528-z

Bureau of Enquiry and Analysis for Civil Aviation Safety. (2012). *Final Report AF447*. Retrieved from https://www.bea.aero/docspa/2009/f-cp090601.en/pdf/f-cp090601.en.pdf

Carlsen, A. N., Chua, R., Inglis, J. T., Sanderson, D. J., & Franks, I. M. (2008). Motor preparation in an anticipation-timing task. *Experimental Brain Research*, *190*, 453-461. https://doi.org/10.1152/jn.00878.2007

Casner, S. M., Geven, R. W., Recker, M. P., & Schooler, J. W. (2014). The retention of manual flying skills in the automated cockpit. *Human Factors*, *56*, 1506–1516. https://doi.org/10.1177/0018720814535628

Casner, S. M., Geven, R. W., & Williams, K. T. (2013). The effectiveness of airline pilot training for abnormal events. *Human Factors*, *55*, 477–485. https://doi.org/10.1177/00187208124668

Davis, M. (1992). The role of the amygdala in conditioned fear. In J. P. Aggleton (Ed.), *The amygdala: Neurobiological aspects of emotion, memory, and mental dysfunction* (pp. 255–306). New York, NY: Wiley-Liss.

de Boer, R. J. & Hurts, K. (2017). Automation surprise: Results of a field survey of Dutch pilots. *Aviation Psychology and Applied Human Factors*, *7*, 28–41. http://dx.doi.org/10.1027/2192-0923/a000113

Dixon, S. R., & Wickens, C. D. (2006). Automation reliability in unmanned aerial vehicle control: A reliance-compliance model of automation dependence in high workload. *Human Factors, 48*, 474–486. https://doi.org/10.1518/001872006778606822

Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007). On the independence of compliance and reliance: Are automation false alarms worse than misses? *Human Factors*, *49*, 564–572. https://doi.org/10.1518/001872007X215656

Dutch Safety Board. (2010). Final Report TA1951. Retrieved from https://catsr.vse.gmu.edu/SYST460/TA1951_AccidentReport.pdf

Ekman, P., Friesen, W. V., & Simons, R. C. (1985). Is the startle reaction an emotion? *Journal Of Personality And Social Psychology*, *49*, 1416-1426. doi:10.1037/0022-3514.49.5.1416

European Aviation Safety Agency (2013). *EASA Automation Policy: Bridging Design and Training Principles.* Retrieved from https://www.easa.europa.eu/sites/default/files/dfu/sms-docs-EASp-SYS5.6---Automation-Policy---14-Jan-2013.pdf

European Aviation Safety Agency (2018). *Crew resource management (CRM) training.* Notice of Proposed Amendment 2014-17.

Federal Aviation Administration (2013). *Airline Transport Pilot Certification Training Program* (FAA-AC: 61-138). Washington, DC: U.S. Department of Transportation. Retrieved from https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_61-138.pdf

Federal Aviation Administration (2014). *Flight Simulation Training Device Qualification Standards for Extended Envelope and Adverse Weather Event Training Tasks*. In Aviation Rulemaking Advisory Committee (FAA)(Ed.) (Vol. 79, pp. 39462-39752): Federal Register

Federal Aviation Administration (2015). *Stall Prevention and Recovery Training*. (FAA-AC: 120-109A). Washington, DC: U.S. Department of Transportation. Retrieved from https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_120-109A.pdf

Federal Aviation Administration (2017). *Upset Prevention and Recovery Training* (FAA-AC: 120-111). Washington, DC: U.S. Department of Transportation. Retrieved from https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_120-111_CHG_1.pdf

Federal Aviation Administration (FAA) & Aviation Supplies & Academics (ASA). (2015). *FAR/AIM 2016 (PDF eBook): Federal aviation Regulations/Aeronautical information manual.* Newcastle: Aviation Supplies and Academics, Inc.

Fernandes, A., & Braarud, P. Ø. (2015). Exploring measures of workload, situation awareness and task performance in the main control room. *Procedica Manufacturing*, *3*, (1281 -1288). doi.org/10.1016/j.promfg.2015.07.273

Fetcho, J. R., & McLean, D. L. (2010). Startle response. In L. R. Squire (Eds), *Encyclopedia of Neuroscience* (pp. 375-379). https://doi.org/10.1016/B978-008045046-9.01973-2

Foster, M.I., & Keane, M.T. (2015). Why some surprises are more surprising than others: Surprise as a metacognitive sense of explanatory difficulty. *Cognitive Psychology*, 81, 74-116. https://doi.org/10.1016/j.cogpsych.2015.08.004

Geels-Blair, K., Rice, S., & Schwark, J. (2013). Using system-wide trust theory to reveal the contagion effects of automation false alarms and misses on compliance and reliance in a simulated aviation task. *International Journal of Aviation Psychology*, *23*, 245–266. https://doi.org/10.1080/10508414.2013.799355

Gillen, M. W. (2016). *A study evaluating if targeted training for startle effect can improve pilot reactions in handling unexpected situations in a flight simulator* (Doctoral dissertation). Retrieved from http://huntlibrary.erau.edu/

Harris, D. P. (2011). *Human performance on the flight deck*. Retrieved from https://ebookcentral.proquest.com

Haslbeck, A., & Hoermann, H. J. (2016). Flying the needles. *Human Factors*, *58*, 533–545. https://doi.org/10.1177/0018720816640394

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence Based Nursing, 18*, 66. https://ebn.bmj.com/content/ebnurs/18/3/66.full.pdf

Hew, P. C. (2017). Detecting occurrences of the “substitution myth”: A systems engineering template for modeling the supervision of automation. *Journal of Cognitive Engineering and Decision Making*, *11*, 184–199. https://doi.org/10.1177/1555343416674422

Hilscher, M. B., Breiter, E. G., & Kochan, J. a. (2012). *From the couch to the cockpit: Psychological considerations during high-performance flight training*, 1–11. Retrieved from https://www.apstraining.com/wp-content/uploads/Psychological-Considerations-During-High-Performance-Flight-Training-2005-Hilscher-Breiter-Kochan.pdf

Horstmann, G. (2006). Latency and duration of the action interruption in surprise. *Cognition & Emotion*, 20, 242-273. https://doi:10.1080/02699930500262878

International Air Transport Association. (2015a). *Guidance Material and Best Practices for the Implementation of Upset Prevention and Recovery Training*. Retrieved from https://www.iata.org/whatwedo/ops-infra/training-licensing/Documents/gmbp_uprt.pdf

International Air Transport Association. (2015b). *Loss of Control In-Flight Accident Analysis Report 2010-2014.* Retrieved from https://www.iata.org/whatwedo/safety/Documents/LOC-I-1st-Ed-2015.pdf

Khemka, S., Tzovara, A., Gerster, S., Quednow, B. B., & Bach, D. R. (2017). Modeling startle eyeblink electromyogram to assess fear learning: Modeling startle-blink EMG to assess fear learning. *Psychophysiology*, *54*(2), 204-214. doi:10.1111/psyp.12775

Koch, M. (1999). The neurobiology of startle. *Progress in Neurobiology*, *59*, 107–128. https://doi.org/10.1016/S0301-0082(98)00098-7

Kochan, J. A., Breiter, E. G., & Jentsch, F. (2004). Surprise and unexpectedness in flying: Data base reviews and analyses. *In Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (pp. 335–339). Santa Monica, CA: Human Factors and Ergonomics Society. https://doi.org/10.1177/154193120404800313

Landman, A., Groen, E. L., van Paassen, M. M. (René)., Bronkhorst, A. W., & Mulder, M. (2017a). Dealing With unexpected events on the flight deck: A conceptual model of startle and surprise. *Human Factors*, *59*(8), 1161-1172. doi:10.1016/S0301-0082(98)00098-7

Landman, A., Groen, E. L., van Paassen, M. M. (René), Bronkhorst, A. W., & Mulder, M. (2017b). The influence of surprise on upset recovery performance in airline pilots. *The International Journal of Aerospace Psychology*, *27*, 2-14. https://doi.org/10.1080/10508414.2017.1365610

Landman, H. M., van Oorschot, P., van Paassen, M. M., Groen, E. L., Bronkhorst, A. W., & Mulder, M. (2018). Training pilots for unexpected events: A simulator study on the advantage of unpredictable and variable scenarios. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *60(*6), 793-805. doi:10.1177/0018720818779928

Lee, J. D. (2008). Review of a pivotal human factors article: "Humans and Automation: Use, Misuse, Disuse, Abuse." *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 404–410. https://doi.org/10.1518/001872008X288547

Martin, W. L., Murray, P. S., & Bates, P. R. (2012). The effects of startle on pilots during critical events: A case study analysis. *Proceedings of 30th EAAP Conference: Aviation Psychology & Applied Human Factors.* 387–394. Retrieved from https://research-repository.griffith.edu.au/bitstream/handle/10072/54072/82496_1.pdf?sequence=1

Martin, W. L., Murray, P. S., Bates, P. R., & Lee, P. S. Y. (2015). Fear-potentiated startle: A review from an aviation perspective. *International Journal of Aviation Psychology*, *25*(2), 97–107. https://doi.org/10.1080/10508414.2015.1128293

Martin, W. L., Murray, P. S., Bates, P. R., & Lee, P. S. Y. (2016). A flight simulator study of the impairment effects of startle on pilots during unexpected critical events. *Aviation Psychology and Applied Human Factors, 6*, 24–32. https://doi.org/10.1027/2192-0923/a000092

Moriarty, D. C. (2015). *Practical human factors for pilots*. Retrieved from https://ebookcentral.proquest.com

NASA Ames Research Centre (1986). National Aeronautics and Space Administration (NASA). Task Load Index (TLX). Retrieved from https://humansystems.arc.nasa.gov/groups/TLX/downloads/TLX.pdf

National Transportation Safety Board. (2010). *Aircraft Accident Report* (NTSB/AAR-10/01). Retrieved from https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR1001.pdf

Parasuraman, R., & Manzey, D.H. (2010). Complacency and bias in human use of automation: An attentional integration. *The Journal of the Human Factors and Ergonomics Society 52*. 381–410. doi:10.1177/0018720810376055

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*, 230. https://doi.org/10.1518/001872097778543886

PARC/CASR Flight Deck Automation. (2013). Operational use of flight path management system. Retrieved from https://www.faa.gov/aircraft/air_cert/design_approvals/human_factors/media/OUFPMS_Report.pdf

Ponce, P., Polasko, K., & Molina, A. (2016). Technology transfer motivation analysis based on fuzzy type 2 signal detection theory. *AI & Society, 31*, 245-257. https://doi.org/10.1007/s00146-015-0583-x

Rankin, A., Woltjer, R., Field, J. (2016). Sensemaking following surprise in the cockpit—a re-framing problem. *Cognition, Technology & Work*, *18*(4), 623-642. doi:10.1007/s10111-016-0390-2

Rivera, J., Talone, A. B., Boesser, C. T., Jentsch, F., & Yeh, M. (2014). Startle and surprise on the flight deck: Similarities, differences, and prevalence. *Proceedings of the Human Factors and Ergonomics Society*, 2014–January, 1047–1051. https://doi.org/10.1177/1541931214581219

Rogers, R. O. (2007). Preliminary results of an experiment to evaluate transfer of low-cost , simulator-based airplane upset-recovery training. *Aerospace \\ Federal Aviation Administration \\ Embry-Riddle Aeronautical University*, (October), 01–24. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/a475565.pdf

Sauer, J., Chavaillaz, A., & Wastell, D. (2016). Experience of automation failures in training: effects on trust, automation bias, complacency and performance. *Ergonomics*, *59*, 767–780. https://doi.org/10.1080/00140139.2015.1094577

Schuman, D. L., & Killian, M. O. (2019). Pilot study of a single session heart rate variability biofeedback intervention on veterans' posttraumatic stress symptoms. *Applied Psychophysiology and Biofeedback*, *44*(1), 9-20. doi:10.1007/s10484-018-9415-3

Schützwohl, A. (1998). Surprise and schema strength. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1182-1199. doi:10.1037/0278-7393.24.5.1182

Simons, R. C. (1996). Boo!: Culture, experience, and the startle reflex. Oxford: Oxford University Press.

Suppiah, S. (2019). *Impact of Electronic Flight Bag on Pilot Workload* (Graduate Capstone Project). Retrieved from https://commons.erau.edu/student-works/141

Thackray, R. I. (1988). Performance recovery following startle: A laboratory approach to the study of behavioral response to sudden aircraft emergencies (No. DOT/FAA-AM-88/4). Washington, DC: Federal Aviation Administration, Office of Aviation Medicine.

Thackray, R., & Touchstone, R. M. (1970). Recovery of motor performance following startle. *Perceptual Motor Skills*, *30*, 279-292. https://doi.org/10.2466/pms.1970.30.1.279

Thackray, R. I., & Touchstone, R. M. (1983). Rate of initial recovery and subsequent radar monitoring performance following a simulated emergency involving startle (Report No.FAA-AM-83-13). Washington, DC: Federal Aviation Administration.

Yeomans, J. S., Li, L., Scott, B. W., & Frankland, P. W. (2002). Tactile, acoustic and vestibular systems sum to elicit the startle reflex. *Neuroscience and Biobehavioral Reviews*, *26*(1), 1-11. Retrieved from http://hsinnamon.web.wesleyan.edu/wescourses/NSB-Psyc255/Readings/11.%20Defense%20Circuitry/Yeomans.pdf

Wallis, T. S. A., & Horswill, M. S. (2007). Using fuzzy signal detection theory to determine why experienced and trained drivers respond faster than novices in a hazard perception test. *Accident Analysis and Prevention*, *39*, 1177–1185. https://doi.org/10.1016/j.aap.2007.03.003

Wickens, C. D., Clegg, B. A., Vieane, A. Z., & Sebok, A. L. (2015). Complacency and automation bias in the use of imperfect automation. *Human Factors*, *57*, 728–739. https://doi.org/10.1177/0018720815581940

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). Engineering psychology and human performance (Fourth ed.). Boston: Pearson

Wiggins, M. W., Stevens, C., Chidester, M. T. R., Hayward, M. B. J., Johnston, C. N., & Dismukes, D. R. K. (2016). Aviation social science: Research methods in practice (1st ed.). London: Routledge Ltd. doi:10.4324/978131526

# Appendix A

## Permission to Conduct Research

**Embry-Riddle Aeronautical University**
**Application for IRB Approval**
**Limited or Expedited Determination**

**Principal Investigator:** Rahim Daud Agha

**Other Investigators:** Dr. Jennifer Thropp, Dr. Andrew R. Dattel

**Role:** Student **Campus:** Daytona Beach **College:** Aviation/Aeronautics

**Project Title:** A flight Simulator Study to Propose and Analyze Scenarios that can Induce Startle and Surprise in General Aviation Pilots

**Review Board Use Only**

**Initial Reviewer:** Teri Gabriel **Date:** 10/02/2019 **Approval #:** 20-038

**Exempt:** No

IRB Member Reviewer Signature: Cheryl L. Marcham, PhD, CIH, CSP, CHMM, FAIHA Digitally signed by Cheryl L. Marcham, PhD, CIH, CSP, CHMM, FAIHA Date: 2019.10.08 18:43:59 -05'00' Date: 10/08/2019

Dr. Michael Wiggins IRB Chair Signature: Michael E. Wiggins, Ed.D. **Date:** 10/09/2019

**Brief Description:**

The goal of this study is to propose scenarios that can trigger a startle and surprise reaction from the pilots. A further goal is to evaluate those scenarios to determine how much pilot performance deteriorates during those events. Participants will be asked to fly six flight scenarios in the Advance Flight Simulation Center.

This research falls under the expedited category as per 45 CFR 46.110 (b) because one of the following apply:

(1) ☐ some or all of the research appears on the list provided by the Office of Human Research Protections and/or are found by the reviewer(s) to involve no more than minimal risk;

(2) ☐ minor changes in previously approved research during the period for which approval is authorized;

(3) ☐ research for which **Limited IRB review** is a condition of Exemption;

[Under an expedited review procedure, the review may be carried out by the IRB chairperson or by one or more experienced reviewers designated by the chairperson from among members of the IRB. In reviewing the research, the reviewers may exercise all of the authorities of the IRB except that the reviewers may not disapprove the research. A research activity may be disapproved only after review in accordance with the nonexpedited procedure set forth in §46.108(b).]

a. ☑ Prospective collection of biological specimens for research purposes by noninvasive means.

b. ☐ Research on individual or group characteristics or behavior (including, but not limited to, research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies. (NOTE: Some research in this category may be exempt from the HHS regulations for the protection of human subjects §46.101(b)(2) and (b)(3). This listing refers only to research that is not exempt.) [This means research that presents *more than minimal risk to human subjects*.]

(3) Research for which **limited** IRB review is a condition of exemption as follows:

a. ☐ Research that only includes interactions involving educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of public behavior (including visual or auditory recording) if the information obtained is recorded by the investigator in such a manner that the identity of the human subjects can readily be ascertained, directly or through identifiers linked to the subjects. §46.104(d)(2)(iii)

b. ☐ Research involving benign behavioral interventions in conjunction with the collection of information from an adult subject through verbal or written responses (including data entry) or audiovisual recording if the subject prospectively agrees to the intervention and information collection and the information obtained is recorded by the investigator in such a manner that the identity of the human subjects can readily be ascertained, directly or through identifiers linked to the subjects. §46.104(d)(3)(i)(C)

c. ☐ Storage or maintenance for secondary research for which broad consent is required: Storage or maintenance of identifiable private information or identifiable biospecimens for potential secondary research use. §46.104(d)(7)

d. ☐ Secondary research for which broad consent is required: Research involving the use of identifiable private information or identifiable biospecimens for secondary research use, if the following criteria are met:

(i) Broad consent for the storage, maintenance, and secondary research use of the identifiable private information or identifiable biospecimens was obtained.

(ii) Documentation of informed consent or waiver of documentation of consent was obtained. §46.104(d)(8)

# Appendix B

## Informed Consent Form

**INFORMED CONSENT FORM**

A FLIGHT SIMULATOR STUDY TO MEASURE PERFORMANCE AND VITAL SIGNS WHILE LANDING AN AIRCRAFT

**Purpose of this Research:** I am asking you to take part in a research project for the purpose of collecting data to develop a baseline on how heart rate, blood pressure, and respiration rate changes while landing an aircraft. You will be asked to fly six flight scenarios using the Elite-P1 135 AATD with X-Plane Flight Simulator. To complete the flight scenarios, the researcher will provide you with aeronautical chart in digital form which you can view using an iPad which will also be provided to you. The study will be conducted in the Advance Flight Sim Center Room #101. The study will take less than an hour to finish.

**Eligibility:** To be eligible for this study, you must be enrolled in college, at least 18 years of age and have a multiengine license.

**Risks or discomforts:** The risks of participating in this study are minimal, no greater than playing a desktop video game. General risks such as motion sickness is unlikely as the flight simulator is a non-motion desktop simulator. The risk of motion sickness includes eye strain, headaches, nausea, sweating, burping, or dry mouth. If you experience any discomfort, you may stop and rest, or just stop the simulation. You will always be in the view of the researcher throughout the study. Fingertip oximeter and respiration rate belt is cleaned using alcohol wipes.

**Benefits:** While there are no benefits to you as a participant, your participation in this research may help us understand how health vitals change while flying.

**Confidentiality of records:** Your individual information will be protected in all data resulting from this study. No personal information will be collected other than basic demographic descriptors. All responses will be kept in a password protected file on a password protected computer. All information about participants who have successfully completed the study will be completely deleted at the end of the summer 2022 semester. Participants who withdrawal or opt out, there data will be deleted immediately and will not be used for this study. No video or audio recording will be done during the experiment. Information collected as part of this research ***will not be used*** or ***distributed*** for future research studies.

**Compensation:** You will be compensated $20 for participating in this study. If you begin the study and decide to discontinue during the study, you will not be compensated.

**Contact:** If you have any questions or would like additional information about this study, please contact Rahim Agha, aghar@my.erau.edu or the faculty member overseeing this project, Dr. Jennifer E.

Thropp, throppj@erau.edu. For any concerns or questions as a participant in this research, contact the Institutional Review Board (IRB) at 386-226-7179 or via email teri.gabriel@erau.edu.

**Voluntary Participation:** Your participation in this study is completely voluntary. You may discontinue your participation at any time without penalty or loss of benefits to which you are otherwise entitled. Should you wish to discontinue the research at any time, you data will be immediately deleted. No personal information attached to the data will be used.

1) **Participant Privacy:** Any personal information that can identify you will be removed from the data collected and this data will **not** be used or distributed for ***future research studies***.

**CONSENT.** By signing below, I certify that I am a college student, at least 18 years of age and have a multiengine license. I further verify that I understand the information on this form, that the researcher has answered any and all questions I have about this study, and I voluntarily agree to participate in the study. I will not disclose any information pertaining to this study to anyone.

Signature of Participant ________________ Date: ________________

Printed Name of Participant ________________

# Appendix C

## NASA-TLX and Participant Perception Question

## NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

| Name | Task | Date |
|---|---|---|
| | | |

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

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

Very Low — Very High

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

Perfect Failure

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

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed were you?

Very Low — Very High

Which options best describes your perception of the task? (Check all that apply)

☐ Surprising (You were not expecting it)

☐ Startling (You were shocked, alarmed, scared, frightened, or disturbed)

☐ Confusing

☐ Unsettling

☐ None of the above