12-1-2016

A Study Evaluating if Targeted Training for Startle Effect can Improve Pilot Reactions in Handling Unexpected Situations in a Flight Simulator

Michael Gillen

Recommended Citation

Gillen, Michael, "A Study Evaluating if Targeted Training for Startle Effect can Improve Pilot Reactions in Handling Unexpected Situations in a Flight Simulator" (2016). *Theses and Dissertations*. 345.
https://commons.und.edu/theses/345

A STUDY EVALUATING IF TARGETED TRAINING FOR STARTLE EFFECT CAN IMPROVE PILOT REACTIONS IN HANDLING UNEXPECTED SITUATIONS IN A FLIGHT SIMULATOR

By

Michael William Gillen
Bachelor of Science, University of North Dakota, 1992
Master of Science, University of North Dakota, 2008

A Dissertation

Submitted to the Graduate Faculty
of the
University of North Dakota
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Grand Forks, North Dakota
December
2016

This dissertation, submitted by Michael William Gillen in partial fulfillment of the requirements for the Degree of Doctor of Philosophy from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

Chairperson – Dr. Warren Jensen

Mr. James Higgins

Cheryl Hunter

Dr. Cheryl Hunter

Dr. Kimberly Kenville

Dr. Vadim Rygalov

This dissertation meets the standards for appearance, conforms to the style and format requirements of the Graduate School of the University of North Dakota, and is hereby approved.

Dr. Grant McGimpsey

September 28, 2016

Date

PERMISSION

Title	A Study Evaluating if Targeted Training for Startle Effect Can Improve Pilot Reactions in Handling Unexpected Situations in a Flight Simulator.
Department	Aviation
Degree	Doctor of Philosophy

In presenting this dissertation in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my dissertation work or, in his absence, by the Chairperson of the department or the Dean of the Graduate School. It is understood that any copying or publication or other use of this dissertation or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Michael W. Gillen

September 20, 2016

TABLE OF CONTENTS

CHAPTER

LIST OF FIGURES

LIST OF TABLES

Table Page

ACKNOWLEDGMENTS

There are so many people who helped me along this journey and it would be impossible to name them all. I would like to acknowledge a few special people whom without their support this work would have never been completed. I would like to thank my family including my parents for always supporting me in this effort. Their words of encouragement kept me going when the coursework seemed daunting. I also would like thank my committee who always offered sound advice and guidance developing and writing this dissertation. Their patience and willingness to offer constructive criticism and suggestions were invaluable. Finally, I would like to acknowledge two individuals who saw potential in a young aviator and gave me sound advice and a few breaks along the way. With heartfelt thanks I acknowledge Captain Jim Corley (TWA) and Mr. Elmer Schaal. These two individuals are no longer with us, however, their ideals and philosophy has helped carry me through my academic endeavors and my career as a professional aviator.

DEDICATION

It would have been nearly impossible to take on a work of this magnitude without the loving support of my wife Sonya Gillen. This work is dedicated to her for her unfailing love, guidance, and understanding throughout this long endeavor.

ABSTRACT

Recent airline accidents point to a crew's failure to make correct and timely decisions following a sudden and unusual event that startled the crew. This study sought to determine if targeted training could augment decision making during a startle event. Following a startle event cognitive function is impaired for a short duration of time (30-90 seconds). In aviation, critical decisions are often required to be made during this brief, but critical, time frame.

A total of 40 volunteer crews (80 individual pilots) were solicited from a global U.S. passenger airline. Crews were briefed that they would fly a profile in the simulator but were not made aware of what the profile would entail. The study participants were asked to complete a survey on their background and flying preferences. Every other crew received training on how to handle a startle event. The training consisted of a briefing and simulator practice. Crew members (subjects) were either presented a low altitude or high altitude scenario to fly in a full-flight simulator.

The maneuver scenarios were analyzed using a series of one-way ANOVAs, t-tests and regression for the main effect of training on crew performance. The data indicated that the trained crews flew the maneuver profiles significantly better than the untrained crews and significantly better than the Federal Aviation Administration (FAA) Airline Transport Pilot (ATP) standards. Each scenario’s sub factors were analyzed using regression to examine for specific predictors of performance. The results indicate that in the case of the high altitude profile, problem diagnosis was a significant factor, in the low altitude profile, time management was also a significant factor. These predicators can be useful in further targeting training.

The study’s findings suggest that targeted training can help crews manage a startle event, leading to a potential reduction of inflight loss of control accidents. The training was broad and intended to cover an overall aircraft handling approach rather than being aircraft specific. Inclusion of this type of training by airlines has the potential to better aid crews in handling sudden and unusual events.

CHAPTER I

INTRODUCTION

Emergencies in aircraft often involve high-stress decision-making, which must be accomplished correctly in real time, often with limited information. Crews are often startled at the onset of such events. Even correct decision-making at the outset of an emergency may not guarantee a successful outcome. Unfortunately, incorrect initial decisions at the start of an emergency often result in delayed aircraft recovery and in some cases lead to an undesired aircraft state (UAS). Decisions in stressful environments are often made with information from past experiences, training, and pattern matching (Rasmussen, 1983). Although each emergency is surrounded by unique circumstances, training over a broad array of scenarios and circumstances may give flight crews enough background information to manage the situation for a successful outcome.

A study completed by Woodhead (1969) found decrements on a decision-making following a startle event. Thackray (1969) also found that major performance decrement following a startle event probably occurs within the first few seconds. In the official report on Air France flight 447, the Bureau d'Enquêtes et d'Analyses (BAE, 2012) stated, “The startle effect played a major role in the destabilization of the flight path and in the two pilots understanding the situation.” Startle training may be a key element in effective emergency flight training. During such training, crews are exposed to different, complex, and unusual situations they would not normally encounter under normal flight conditions. The purpose of this type of training is to develop the pattern behavior of systematically dealing with complex emergencies.

Problem Statement

Recent airline accidents point to a rapid degradation from controlled flight following an unusual event when the flight crew becomes startled. There has been very little training among airline crews on how to successfully manage a sudden and often stressful event that requires quick and accurate decision-making (BAE, 2012). Accident data has indicated that when an incorrect decision is made, the likelihood of a successful outcome decreases (Hilscher, Breiter, & Kochan, 2012). This study seeks to determine if specific and targeted training can help mitigate the effects of flight crews being startled by implementing a set of techniques designed to help stabilize the cognitive thought process and bridge the time of cognitive degradation.

Purpose of the Study

The intent of this mixed methods study is to test the theory that enhanced specific training can provide an effective countermeasure to fill the temporary cognitive degradation that occurs during a startle event. The study will use both survey and observed simulator performance data to test the theory. If the hypothesis is correct, specific training could be added to airline qualification programs to better equip airline crews in dealing with sudden, unusual events.

Research Design

The research is designed as a mixed methods study using quantitative analysis methods. In the first phase, a survey will be conducted of the participating air carrier pilots. This survey will be used to gauge the pilot's own perceptions of their flying skills during a startle event. The analysis will explore for common threads of pilot thinking and reactions. The results of the survey will be compared and correlated to the data from the aircraft simulator scenario sets.

The second phase of the study involves evaluating professional airline pilots flying two different scenarios in an FAA approved Level-D full flight simulator (FFS). The scenarios will be flown by a crew consisting of a captain and first officer, similar to what would happen in actual line operations. Each crew will be presented either a low or high altitude scenario depending on the day of the week. Randomly selected crews will receive training on handling the aircraft during a startle event. The pilot group that does not receive the startle training will be considered the control group. This group will be referred to as the untrained group for the purposes of this study. The training consists of both a briefing and simulator practice. The training briefing is via an in person discussion on the proper pitch, power, and bank, settings that should be flown in an unusual event. The briefing also discusses time recognition especially at low altitudes and fuel states. The training continues with simulator practice using the techniques discussed in the briefing. The practice sessions are not the same as the evaluation profiles. Each crew will practice both a low altitude and high altitude scenario.

The first test scenario is a low altitude and low fuel profile. Time pressure and an unexpected missed approach combine to form the startle event and event evaluation begins at the missed approach. The second test scenario is a high altitude profile. The profile induces a loss of air data followed by an engine fire bell that causes the startle event. Evaluation begins at the loss of air data. In terms of procedures, the loss of air data is often referred to by aircraft manufacturers as Mach/Airspeed Unreliable. Data analysis will consist of regression, ANOVA, and post-hoc Tukey tests. The analysis looks for differences between trained and untrained groups with regression looking for differences within groups such as previous experience.

Research Questions

The research questions surround the cognitive gap that is perceived to exist during a startle event and to what extent training can mitigate the gap.

1. Can targeted training be successful in helping pilots maintain aircraft control during an unusual and sudden startle event?
    a. Does the spatial proximity of the event have any effect on the outcome (low or high altitude)?
    b. Since accident data indicates that accidents occur more frequently with the captain flying and on the first day and first leg of a trip, does the pilot flying, either the captain or the first officer, have any effect on the successful outcome of the event?

2. Is there an indication in the pilot's survey answers that is a predictor of being able to successfully handle a startle event?

Assumptions

The list below is not meant to be all-encompassing but to inform the reader as to the major aspects involved in the study. This study develops findings based on the following assumptions.

1. Each participant is a qualified, Airline Transport Pilot (ATP) certified FAR part 121 jet transport pilot employed by a U.S. air carrier.
2. Each participant has spent at least one year in the specific seat (Captain or First Officer) and type of aircraft. It is assumed that after one year of experience on a particular aircraft, that the pilot will be normalized to flying that particular aircraft (the aircraft will not be "new" to them).
3. Each pilot is current and qualified in the respective aircraft. Current and qualified indicates that the pilot can be scheduled to fly a regular passenger trip at any time.
4. Each pilot is considered a line pilot. For the purpose of this assumption, line pilot means that each pilot flies their respective airplane at regular intervals. Line pilots include; Captains, First Officers, Line Check Airman, and Instructors/Evaluators

5. Except for the group that receives training, the pilots have no prior knowledge or practice of the maneuver that is to be flown and is given no opportunity to practice it beforehand.
6. Each pilot is assumed to fly to the best of their ability during the maneuver.

Limitations

The study sought to mitigate possible limitations that could skew the results. Although each study has a set of unique limitations, the results of this study should be considered in the context of the limitations listed below.

1. The maneuver sets are flown in a simulator that is realistic in nature but involves simulation limits specifically the general lack of g-forces.
2. The study only looks at two maneuver sets.
3. Aircraft emergencies are often dynamic and are unique to each situation.
4. The study only observed professional pilots flying transport category aircraft.
5. Crews volunteered for this study which may indicate a higher awareness of safety.
6. Crews were sampled directly after their recurrent training which may increase their proficiency above what might be expected in normal line operations.

Expected Findings

The study has four major groups of subjects that are being compared to a set standard as defined the FAA and airline policy for successful outcomes. The study expects to find crews that receive training will become more aware and proficient on how to handle an inflight emergency with regards to the first few critical decisions and/or actions. The expectations are that the trained group for both low and high altitude scenarios will show a statistically significant increase in performance when compared to the non-trained groups. In addition, when collapsed for maneuvers sets, the trained groups should show a significant increase in performance than the non-trained group. Items such as past experience and current outside flying might be reasonably expected to also influence success.

CHAPTER II

LITERATURE REVIEW

Introduction

A literature review was conducted of pertinent articles related to this study. Although there were no direct studies on this particular problem, there were many articles related to decision making, startle effect, and time critical actions required by flight crews. The review begins by relating a discussion on training of unusual events and why practice is important for improved performance. The literature review then discusses what startle effect is and how it effects cognition, decision making and pilot responses. The review then takes the theoretical discussion and relates it to actual aircraft accidents using the official reports as a background. This literature review is not meant to be all encompassing, but to give the reader a broad overview of the issues surrounding this study.

Training and Unusual Events

Accident reports describe many situations where pilots responded to abnormal events in ways that were unexpected from the way that they were trained (Casner, Geven, & Willliams, 2012). Unfortunately, training and testing of professional airline pilots have become somewhat routine and predictable. In short, the flight crews know what to expect as they see the same maneuvers at each training event.

In a study by Dr. Stephen Casner, (when) pilots were evaluated performing routine training events and unexpected (but similar) ones. Pilot responses to the routine events showed little variability. In contrast, pilot responses in the unexpected maneuvers showed great variability from pilot to pilot (Casner, Geven, & Willliams, 2012). The results of the study showed that most pilots generally experience the same sequence of abnormal events, presented under similar circumstances. This is due to both the airline training environment and the regulatory environment as set forth by the FAA. This training calls into question the extent to which pilots have the ability to respond to abnormal events in actual operations (Casner, Geven, & Willliams, 2012). Casner, Geven, and Williams, suggested that such training can lead to shallow and memorized understandings of problem situations which in turn do not lead to an ability to transfer this training to different encounters in actual operations (Casner, Geven, & Willliams, 2012). The end result of the study was that pilots struggle to recognize unexpected situations with the result of considerably delayed responses (Casner, Geven, & Willliams, 2012).

It is unlikely that training alone can eliminate the element of surprise from unexpected events although skill and experience are known to reduce the occurrence and/or severity (Merk, 2009). Furthermore, for unusual events, pilots would benefit from exposure without the use of automation to enable them to better recognize the situation itself rather than respond to an alert (Wiener, 1985). Finally, the most important step in training is to train abnormal events over a wide array of circumstances and operation parameters (Casner, Geven, & Willliams, 2012).

Safety management tends to focus on prevention of errors and failures. In most failure cases, there are opportunities to recover from the failure through the timely and effective application of countermeasures. The aim of the countermeasures to prevent the negative consequences of the failure (Kranse & van der Schaaf, 2001). In the case of unforeseen failures, human operators play a key role in the application of effective countermeasures. Researchers generally agree that the failure compensation process has three phases:

- Phase One: Detection of the fact that something has gone wrong
- Phase Two: Explanation or localization of the causes
- Phase Three: Correction of the problem through planning and execution of countermeasures

A study by Kranse, and van der Schaaf, (2001) asked the question: how does the failure compensation process work, and what factors influence the process? A failure can be a combination of technical, organizational, or human factors. Detection of the situation is always the first failure compensation phase to occur (Kranse & van der Schaaf, 2001). After the detection phase, time often dictates the next step. The corrective action can be performed immediately (usually the case in aviation) or on a longer term for systemic issues. The study examined 50 reported failures at a chemical plant. The failures were all reported via a voluntary safety reporting system similar to Aviation Safety Action Program (ASAP) reporting systems. The study noted that in most detection and localization phases unplanned actions (not trained) occurred. In the detection phase, 46 out of 50 cases (92%) involved unplanned actions (Kranse & van der

Schaaf, 2001). For failures requiring immediate localization, 100% involved some unplanned actions. Such was the same for events requiring immediate correction, where 80% involved unplanned actions (Kranse & van der Schaaf, 2001). Even in events that were not as time critical, where longer term corrections were sought, 77% experienced unplanned actions (Kranse & van der Schaaf, 2001). Generally, the type and severity of potential consequences were the most practical factor in deciding which recovery paths to take. The Kranse and van der Schaaf study has important implications for any type of aviation training in as much as untrained actions will invariably take place in a recovery from an unusual event.

Practice and Performance

A study that was published in (McKinney & Davis, 2003)researched the effects of deliberate practice on crisis situations. Within professional domains, deliberate practice has been positively correlated with improved performance (McKinney & Davis, 2003). The study examined the question, “do the benefits of deliberate practice create superior performance if part of the task is unpracticed?” (McKinney & Davis, 2003). Researchers reviewed decision-making under crisis conditions using a total decision effectiveness model. The model evaluated both the initial assessment and the actions taken for both practiced and unpracticed maneuvers. Additional studies have indicated that deliberate practice results in automated pattern matching of problems with solutions (Richman, Gobet, Stazewski, & Simon, 1996). Practiced skills allow for a more accurate diagnoses of the situation, and improves both speed of action and the accuracy of recall (Klein G. , 1993). Practice may aid in the cognitive processes through enhancement of higher levels of searching and evaluating. This, in turn, might enhance the ability to extrapolate beyond the presented data and make use of long-term memory items which is richer and more organized than short-term memory (Ericcsson, 1996).

For this study, wholly practiced maneuvers are ones that pilots have deliberately practiced either in flight, or more likely, in the simulator. Each of these maneuvers had a best practice solution, which was reinforced on a regular basis. In contrast a partially practiced maneuver was defined as a specific aircraft malfunction, occurring within a wider flying scenario that was novel or unique, and a maneuver that the pilot could not have practiced (McKinney & Davis, 2003). These maneuvers could include items such as multiple system failures, flight control malfunctions, and unusual failure modes. The data for the study was compiled from 173 U.S. Air Force fighter aircraft mechanical malfunction incidents. In each case, the mishap was classified as either wholly practiced or partially practiced. Three independent panels of experienced Air Force pilots reviewed and rated the mishaps. The action taken by the pilots was rated as effective or ineffective (McKinney & Davis, 2003). Furthermore, a third group of evaluators were asked further define where the failure occurred in cases of ineffective responses. The group sought to define if the ineffective responses were in the decision-making process, or in the selection of action process. To evaluate the research questions, logistic regression analysis was used as a tool for predicting group membership in cases where dependent variables are dichotomous (McKinney & Davis, 2003). In total 83 of the events were characterized as wholly practiced, of which 68 ended with effective decisions and 15 with ineffective decisions (McKinney & Davis, 2003). The study concluded that deliberate practice has a positive effect on crisis decision-making performance (McKinney & Davis, 2003). Increased performance was noted for each decision phase for wholly practiced maneuvers. However, the study also found that no relationship existed where the crisis-flying scenario was unpracticed (McKinney, 2003). The study also noted that deliberate practice within the flying domain was not related to overall decision-making performance (McKinney & Davis, 2003).

Startle Effect

Startle effect is quite different than startle training. The startle response has been well researched and documented over the past 60 years. A startle response happens when the human brain is presented with a situation that completely overwhelms the available cognitive resources needed to effectively mitigate the situation. It has been widely established through psychological research that our ability to regulate our own thoughts and behaviors becomes diminished during an emotional event (Hilscher, Breiter, & Kochan, 2012). This diminished ability is compounded by the reliability of today's modern aircraft, which has created a conditioned expectation of normalcy amongst pilots (Martin, Murray, & Bates, 2012). Research has shown that there are considerable cognitive effects on information processing following a startle event. The results indicate that strong cognitive and dexterous impairment could last for up to 30 seconds following a strong startle (Vlasek, 1969; Woodhead M. M., 1959; Woodhead M. , 1969; Thackray & Touchstone, 1970). A pilot describing an encounter with severe turbulence described the situation as "the constant audible warnings came from far-off, detached from the struggle in the cockpit" (Hilscher, Breiter, & Kochan, 2012). The Bureau d'Enquêtes et d'Analyses (BAE) in the official report on Air France Flight 447 (AF447) accident, the made the following statements:

> “The startle effect played a major role in the destabilization of the flight path and in the two pilots understanding the situation. Initial and recurrent training as delivered today does not promote and test the capacity to react to the unexpected. Indeed, the exercises are repetitive, well known to crews and do not enable skills in resource management to be tested outside of this context.
>
> All of the effort invested in anticipation and predetermination of procedural responses does not exclude the possibility of situations with a ‘fundamental surprise’ for which the current system does not generate the indispensable capacity to react.” (BAE, 2012)

The response of the brain, and the consequent behavior is an amalgamation resulting from past experience and general expertise (Isaac, 2012). Once an unusual situation has been determined to exist, pilots attempt to compare the situation with past experiences through a sequence of pattern matching and decision-making. The outcome often relies on the severity of unusual circumstance and emergency training. In addition, other factors include prior experience and the ability to accept the actual facts of the situation. Discrepancies between perception and the actual aircraft state leads to a breakdown of a pilot’s mental picture, which in turn can lead to a loss of situational awareness (Hilscher, Breiter, & Kochan, 2012). Surprising events can place the pilot into a very high state of arousal that can render them ineffective in complex decision-making tasks (Hilscher, Breiter, & Kochan, 2012). The final response is often a strongly developed behavior with the purpose, in extreme cases, of survival. There are examples in which highly trained crews discarded indications from instruments and flight training after a startle event (Isaac, 2012). This leads to disbelieving what is actually presented to the crew from the aircraft's systems. Once an unusual or emergency situation is presented, a pilot will generally be limited in their response. The response tends to fall into patterns a pilot has seen before, and will also be subjected to several decision-making, behavioral biases. An objective that is not addressed in traditional flight training is behavioral management that promotes progressive functionality under conditions of uncertainty and fear (Hilscher, Breiter, & Kochan, 2012).

Automation Bias and Complacency

Automated decision aids support decision-making in complex environments. As such, automation is assuming increasing control of cognitive flight tasks, such as calculating fuel-efficient routes, navigating, or detecting and diagnosing system malfunctions and abnormalities (Mosier, 1998). These systems are designed to support the human cognitive processing of information to correctly assess a given situation and to respond appropriately (Parasuraman & Manzey, 2010). Automation-induced complacency and bias represent closely linked theoretical concepts that show considerable overlap with respect to underlying processes (Parasuraman & Manzey, 2010). Automation complacency can occur when the automation competes for the pilot's attention in a multiple task load environment (Parasuraman & Manzey, 2010). Although somewhat different but interconnected, automation bias results in making both omission and commission errors when the automated decision aids are not accurate (Parasuraman & Manzey, 2010). These two issues affect both novice and expert pilots and cannot generally be mitigated through training.

Research studies have indicated that automation does not always enhance human activity. In some cases, automation can change behavior patterns in ways that are unintended, and cannot be unanticipated by automation designers (Parasuraman & Manzey, 2010). Automated systems in today's modern aircraft are highly accurate and reliable. The end result is that pilots can develop a premature cognitive commitment regarding the information displayed by the automation and disregard other conflicting information (Parasuraman & Manzey, 2010). Automation bias can lead to incorrect decisions that are not based on a complete analysis of the available information and can compromise performance especially in the case of automation failure.

Automated decision aids are misused for two main reasons. The first reason is that automation generated cues are usually salient, and by design, draw the user's attention (Parasuraman & Manzey, 2010). The second major factor is that users have a cognitive bias to assign greater relevance to automated cues over other sources of data (Parasuraman & Manzey, 2010).

In a study by Layton, Smith, and McCoy (1994), a comparison of electronic flight planning tools was examined. Pilots who were given highly automated flight plans spent less time and effort evaluating alternate plans than groups working with manually developed plans. This result was consistent with the cognitive-miser hypothesis of automation bias (Layton, Smith, & McCoy, 1994). The pilots tended to accept the plan generated by the automation even when it produced less than optimal solutions.

Another study on automation bias sought to quantify the effects of automation over-reliance in modern cockpits. This study pointed out the need for pilots to be able to fly the airplane when the automation does not function correctly. Automation bias refers to omission and commission errors resulting from the use of automated cues as a heuristic replacement for vigilant information seeking and processing (Mosier, Skitka, Heers, & Burdick, 1997). Highly automated cockpits tend to change the way pilots perform tasks and make decisions. Researchers have documented problems in the use of advanced automated systems, including mode misunderstanding, failures to understand automated behavior, confusion or lack of awareness concerning what automated systems are doing, and difficulty tracing the functioning or reasoning process of the automated agent (Billings, 1996). Figure 1 below diagrams both positive and negative feedback loops associated with automation. Each loop or pathway can lead to bias on the part of the pilot.

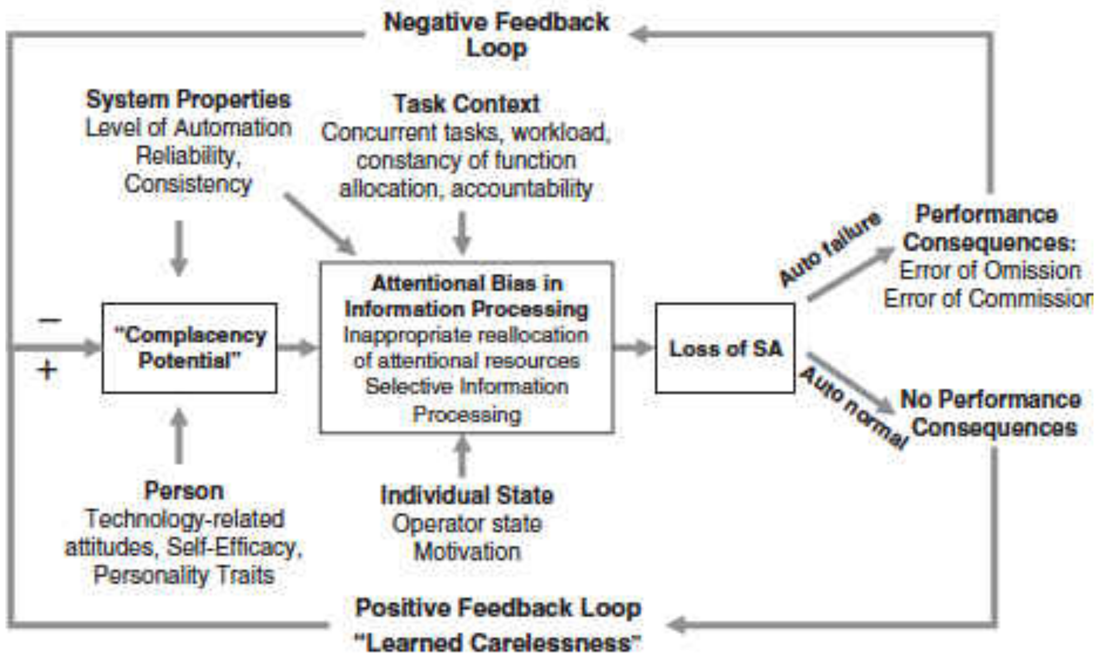


Figure 1. Automation Bias *(Parasuraman & Manzey, 2010)*

Pilots are trained and develop their skills assessment through the use of both system and environmental cues (cross checking of information). In most situations, processing is facilitated by inter-correlations among cues (Wickens & Flach, 1998). In the cross-checking environment, related to older technology aircraft, pilots often looked for many cues in determining if a problem existed. Using these skills, pilots know and look for patterns or combination of cues that are most ecologically valid, reliable, or relevant for diagnosing particular situations. They are able to incorporate contextual information to formulate a workable action plan based on their assessment of these cues (Kaempf & Klein, 1994).

When automated aids are introduced, the pattern of cue utilization is disrupted. Automated aids present powerful and generally highly accurate cues. This leads to the overall attitude that the automated cues are not just another cue, but the most powerful and important cue. These automated decision aids support the general human tendency “to travel the road” of least cognitive effort. People will generally utilize heuristics (cognitive shortcuts) to reduce effort and information load.

For rigid tasks that do not require flexible decision-making automation can often provide the best solution (Cummings, 2016). In time-critical environments that have external and changing constraints, higher levels of automation may not be advisable due to the risks and the complexity of the decision aids not being perfectly stable. (Cummings, 2016). Situation awareness, complacency, and skill degradation are the measurable costs of automation bias.

Breakdown in Coordination

Errors can never be completely eliminated necessitating the need for detection, diagnosis, and recovery (United Airlines, 2016). Event driving tasks and domains have seen a lack of research in error diagnosis and recovery (Nikolic & Sarter, 2007). A study jointly conducted by the Boeing Company and the University of Michigan in 2007sought to some insight into error and disturbance management strategies. The study noted that pilots seldom follow the canonical path to handle disturbance events (Nikolic & Sarter, 2007). A canonical path can be considered the most optimum solution that is technically correct in developing diagnosis and recovery options. Detection of such events were often observed to be delayed due to pilots’ knowledge gaps and time criticality, and in many cases, generic and inefficient recovery strategies were observed (Nikolic & Sarter, 2007). In addition, pilots often relied on high levels of automation to manage the consequences of the induced errors. The study noted that pilots often attempt to diagnose automation-related problems before they responded to the actual disturbance handling (Nikolic & Sarter, 2007), meaning that pilots tended to become focused on the automation instead of flying the aircraft. All 18 pilot crews in the Boeing study struggled at some point with handling events (actual flying) during a simulated flight in the B747. It was noted that the pilots in the study rarely attempted to diagnose the source of the disturbance (Nikolic & Sarter, 2007). The study findings of indicate that poor disturbance management is somewhat related to the design of the automation interfaces (Nikolic & Sarter, 2007).

Cognitive Resources

As skill levels decline, a pilot must devote more cognitive resources when situations such as emergencies, system failures, or other issues that force a pilot into manual flying. In addition, a pilot begins to lose the ability to mentally project where the airplane is in space with regards to altitude, airspeed, and configuration. Simply stated, a pilot's cognitive resources (in fact every human) is finite.

Two basic parameters affect performance: the amount of cognitive resources available to the pilot and the complexity of the task or situation. Task performance depends on the relation between the two parameters, cognitive resources available and the complexity of the situation. As long as the amount of resources consumed by the task is lower than, or equal to, the available amount of memory, task performance will be adequate (Ippel, 1987). However, task performance will gradually decline relative to the degree that tasks impose cognitive loads that exceed the available amount of resources (Ippel, 1987). If too little processing resource is applied (because of limitations to the availability of processing resources), performance failure is to be expected. As more and more resources are applied to the task the likelihood of successful performance increases (Norman & Bobrow, 1975).

Startle Effect and Cognitive Consequences

The startle effect is common to all mammals (Simons, 1996). It consists of an involuntary reaction to an unusual stimulus. This reflex usually happens quickly following the stimulus, generally in as little as 14 milliseconds (Yeomans & Frankland, 1996). Research has suggested a link between common patterns of the startle reflex and the neural pathways involved (Davis, 1986; Eaton, 1984; Landis & Hunt, 1939; Lang, Bradley, & Cuthbert, 1990; LeDoux J. E., 2000; LeDoux J. , 1996; Whalen & Phelps, 2009). These actions involve various senses and muscles and the amygdala in the limbic region of the brain. The initial analysis happens very quickly (500 milliseconds) and results in an aversive reflex away from the stimulus. The startle reaction may last between .3 to 1.5 seconds, depending on the severity (Martin, Murray, & Bates, 2012). An issue arises when the threat persists and the startle reaction becomes a full-blown startle or surprise reaction, otherwise known as “flight or fight.” This process can lead to confusion or delays in processing. When people are startled and the threat persists, such as in a life-threatening aircraft emergency, then the startle reflex is likely to transition into a full startle reaction, with its ensuing activation of the sympathetic nervous system (Martin, Murray, & Bates, 2012). A sudden startling event can have negative effects on performance (Martin, Murray, & Bates, 2012). This is especially detrimental in the case of an emergency were correct decision-making is important to resolve an issue. A study conducted for the FAA (1969), demonstrated that a startle event negatively affected performance and, also noted that recovery of performance following a startle event appears to be quite rapid (Thackray & Touchstone, 1970). Upon initial presentation of the startle stimulus, maximum disruption occurred during the first five seconds after the stimulation, with significant but considerably less disruption after the second 5-second interval lasting from 30 seconds to one minute (Thackray & Touchstone, 1970).

Vlasak (1969) in his study, investigated the effects of startle on a continuous task. This task was measured for accuracy and consistency. Test participants were given a task of continuous mental subtraction. Subtraction was found to be significantly impaired for 15 secconds following stimulation (Vlasek, 1969). For the reaction time tasks, there was insufficient data given to determine the precise duration of impairment, although both were impaired temporarily following startle event (Vlasek, 1969). In a similar study, Woodhead (1969) found decrements on decision-making following sudden noise stimulation that lasted, from 17 to 31 seconds. It would appear from the results of Woodhead’s study, and from others who have investigated performance recovery that major performance decrement following startle probably occurs within the first few seconds (Thackray & Touchstone, 1970). A lesser but significant decrement may last for periods from 10 to 30 seconds after startle. This underperformance has been shown in some accidents to be a period of time where making correct decisions were critical to recovery. Interviews with startled pilots and qualitative data in flight simulator experiments suggest that the negative effects of startle effect are real and in some cases can be significant (Martin, Murray, & Bates, 2012).

Decision-Making Model

There are a number of decision-making models that attempt to explain how pilots make decisions. Aeronautical decision-making is complex and there is not always a clear link between the decisions made and event outcome (Plant & Stanton, 2013). Schema theory explains how people interact and make decisions using stored mental representations, and forms an integral part of the perceptual cycle model (PCM). Aeronautical decision-making is a form of naturalistic decision-making (NDM); (Klein, Calderwood, & Macgregor, 1989) in which decision makers have domain expertise and make decisions in contexts, which are usually characterized by limited time, goal conflicts and dynamic conditions (Plant & Stanton, 2013). A high proportion of pilot errors are related to decisional errors (Diehl, 1991; Orasanu & Martin, 1998; Shappell & Wiegmann, 2009).

Naturalistic decision making (NDM) is complex due in part to the weakly correlated link between event outcome and the decision process. Outcomes cannot always be used as a reliable means to quantify a reasonable decision (Orasanu & Martin, 1998). The perceptual cycle model (Neisser, 1976) is based upon the idea of a reciprocal and cyclical relationship between the operator and the environment (Plant & Stanton, 2013). Neisser presented the view that human thought is closely connected with a person's interaction in the world, both informing the other (Neisser, 1976).

World knowledge (schemata) leads to the anticipation of certain types of information or clues. Accurate cue perception is critical to decision making. In most operational environments, there are multiple cues available; however, when pilots become startled, there is a tendency to reduce the number of cues that are sampled (Wickens & Flach, 1998). Selective cue sampling can lead to a cycle of confusion that further complicates the situation (Hilscher, Breiter, & Kochan, 2012). Schemata can be conceptualized as having mental 'slots' that are used to structure the information linked to them. Schemata represent linked neurons and memories of abstract concepts. They are generally formed from specific instances and allow abstract knowledge to be derived at the time of retrieval by sampling from domain-specific instances (Plant & Stanton, 2013). Schemata are internal knowledge structures that are based on similar experiences that capture the common features of this experience (Lieberman, 2012). The use of schemata in decision-making is advantageous; they act as natural standard operating procedures (SOPs) to direct decision makers to make appropriate responses to environmental stimuli based on previously successful experiences (Plant & Stanton, 2013).

According to the perceptual cycle model (PCM), when an environmental experience is encountered, relevant experiences (schemata) are retrieved to help develop an appropriate response (Plant & Stanton, 2013). This leads to seeking out certain types of additional information in as a way of interpreting that information using a form of bottom-up processing. The environmental experience can result in the modification and updating of cognitive schemata and thus, in turn, influence further interaction with the environment (Plant & Stanton, 2013).

Smith and Hancock (1995) have argued that the usefulness of the PCM explanation lies in the interaction between operator and environment, rather than considering the two separately. Decision issues arise when the selected schema (stored and cataloged memories) is inappropriate for the current situation. In general, pilots were found to utilize a number of different schemas in determining an initial response to a situation. The use of schema aids perception and decision-making (Plant & Stanton, 2013). When the unusual happens, pilots tend to pay closer attention to information related to specific cues relating to the unusual situation instead of seeking out additional information to keep the "big-picture" in mind (Hilscher, Breiter, & Kochan, 2012). Alternative scenario interpretations are usually only considered when they are consistent with preexisting expectations (Muthard & Wickens, 2002).

Morris and Leung (2006) found that mental workload was not significantly increased, when task demand increased if pilots could revert to pre-existing schemata. When inappropriate schemata are selected, incorrect actions and decisions can follow. Over-reliance on pre-existing but inappropriate schemata have been shown to lead to fixation on certain cues in relation to other cues (Stanton, et al., 2010; Plant & Stanton, 2013).

Dual-Process Account of Decision Behavior

In unfamiliar situations, when proven rules are not available, behavior may become goal-controlled using knowledge-based reasoning (Rasmussen, 1983). Coping with complexity is largely due to the availability of a large repertoire of different mental representations of the environment from which rules can be generated ad hoc (Rasmussen, 1983). Purposeful behavior is based on a pilot’s perception of an event and is experiential knowledge of similar situations.

Human behavior can be characterized by three levels of constraints or performance levels. The levels make use of pattern matching and are defined as skill-based, rule-based and knowledge-based performance. Skill based behavior is characterized by sensor-motor performance during activities following a state of intention and generally take place without conscious thought. They are usually smooth, automated and highly integrated (Rasmussen, 1983). This mode is mostly used for quick and accurate movements. The body acts as a multivariable continuous control system synchronizing movements with the behavior of the environment (Rasmussen, 1983). When asked, pilots cannot generally describe their thought process involved in this type of cognition. They refer to it as an automatic-like response. This type of cognition is sought by training departments in response to time critical aircraft emergencies such as an engine failure at rotation where a quick, automated, and precise response is needed.

At the rule based level, information is typically perceived as signs, which serve to activate or modify predetermined actions or manipulations (Rasmussen, 1983). The boundary between skill based and rule based performance is not always distinct, and depends on the level of training and attention of the individual (Rasmussen, 1983). These signs are used to select or modify the rules controlling the sequencing of skilled sub-routines, and cannot be used for functional reasoning to generate new rules.

During unfamiliar situations that have no known rules for control, performance moves to the highest cognition level that is knowledge based (Rasmussen, 1983). In this situation, the goal is formulated based on an analysis of the environment. This mode can be characterized by evaluation of different solutions and can also include trial and error. Figure 2 describes the various levels of how Rasmussen describes his behavior model.

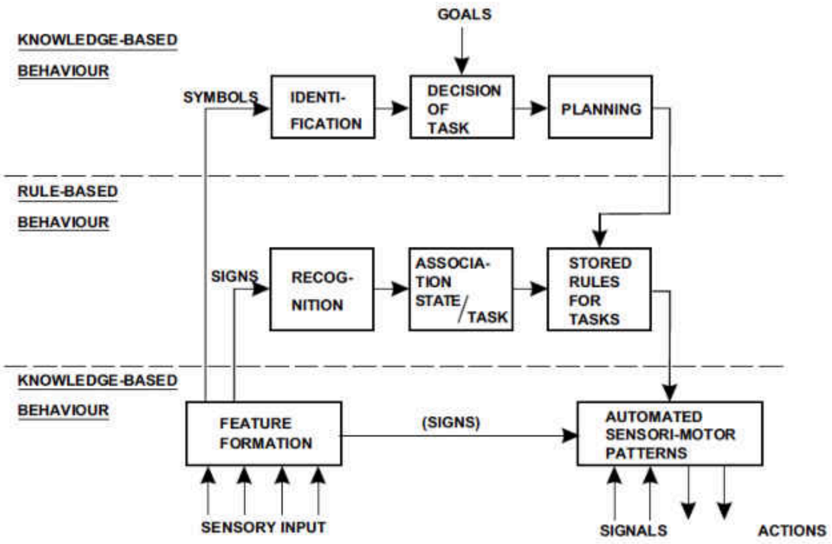


Figure 2. Behavior Model *(Rasmussen, 1983)*

Many decision-process models mark the first step of assessing the situation by observing information and data scanning (Salmon, et al., 2008). The second part of the process involves examining possible solutions depending on the interpretation of this assessment. Decision errors occur when there is a lack of consideration of important data displays. The perceptual step builds the foundational level of Endsley's concept of situation awareness (Endsley, 2006).

Cognition and Emotion

Emotions are evolved situation responses that have multiple aspects. They involve subjective feelings, cognition, information processing, expressive behavior, motivation, and physiological responses (Diamond & Aspinwill, 2003). In fact, cognition and emotion are intertwined constructs (Hilscher, Breiter, & Kochan, 2012). Cognitions that pilots have stored in memory may not be sufficient for exceptional events (Hilscher, Breiter, & Kochan, 2012). Insufficient cognitions changes pilot perception and as a result, place more emphasis on how pilots perceive and interpret events based on their motivational and behavioral significance (Compton, et al, 2003). Pressures such as emotional pressures can alter rational reasoning by shifting decision-making criteria from safety rules to subjective ones (Causse, Dehais, Peran, & Pastor, 2013). Emotion and stress can bias decision-making and cognitive functioning particularly during complex tasks that involve higher cognitive abilities (Causse, Dehais, Peran, & Pastor, 2013). Adding to this issue is an ingrained confidence on the aircraft's reliability. This sense of safety can leave pilots unprepared for sudden emergencies (Hilscher, Breiter, & Kochan, 2012).

Inflight Loss of Control

In terms of aircraft accidents, people often incorrectly associate takeoff and landing phases to be the area where the highest risk occurs. In total numbers of accidents and incidents, as defined by the National Transportation Safety Board (NTSB), takeoffs and landings are the largest accident category. However, according to Boeing (2012), inflight loss of control is the single largest category of fatalities over the past ten years accounting for 1413 fatalities from 18 accidents (Boeing, 2012). Inflight loss of control accidents have more fatalities that both controlled flight into terrain (CFIT) and landing accidents. Many of these inflight loss of control accidents were the result of an unusual event at the beginning of the accident sequence. Loss of control in flight can develop rapidly and suddenly following inappropriate decisions by the flight crew. Figure 11 below shows the various accident fatalities ranked by category with inflight loss of control having the most.

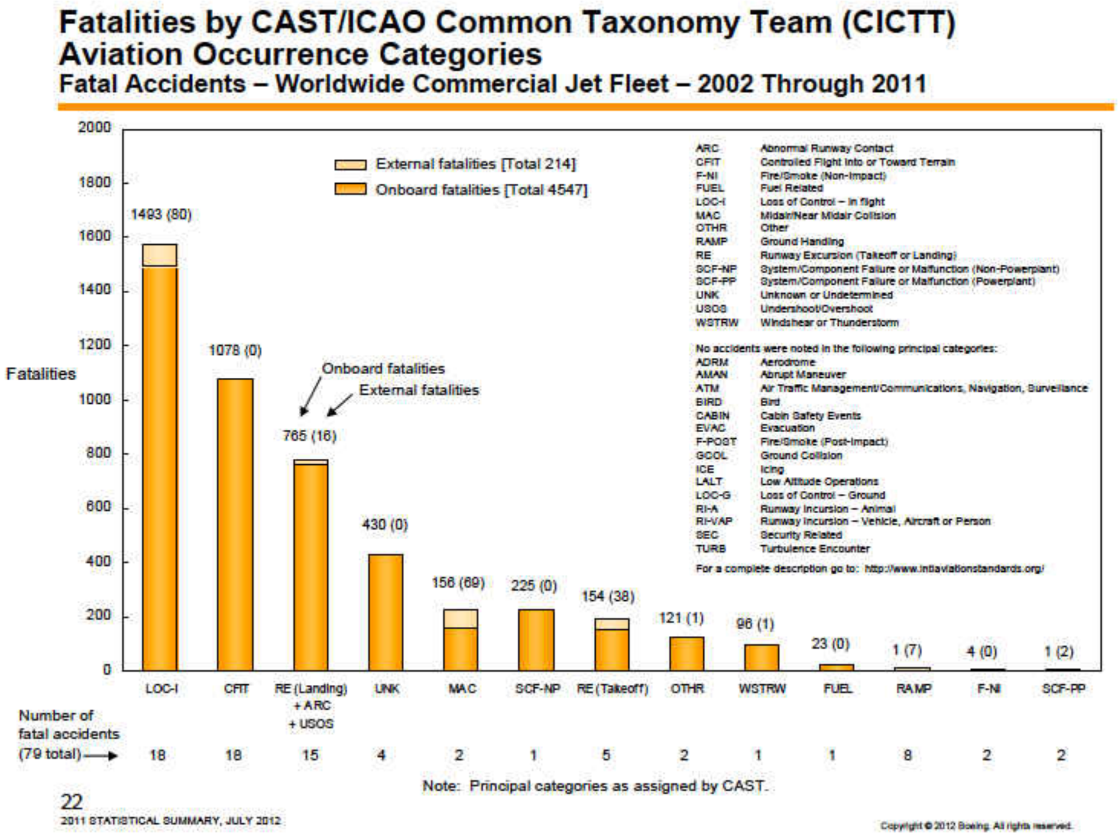


Figure 3. Causes of Accidents *(Boeing, 2012)*

The next sections of this literature review discusses two widely known airline accidents in which inflight loss of control occurred following a startle event.

Aircraft Accident Colgan 3407

As referenced in the last paragraph, inflight loss of control represents a majority of airline accident fatalities. Many of these incidents have been preceded by a startle event. When airline crews are presented with a sudden onset of unusual circumstances, they sometimes react contrary to what are generally accepted correct procedures (NTSB, 2010).

On February 12, 2009, a Colgan Air Bombardier DHC-8-400 (Q400) operating as Continental Connection 3407 crashed while on approach to the Buffalo International Airport. All 45 passengers, 4 crew members, and 1 person on the ground perished as a result of the crash. The aircraft impacted a residential area approximately five nautical miles northeast of the airport while attempting an instrument approach. At the time of the accident, night visual meteorological conditions (VMC) prevailed at the time (NTSB, 2010).

> The METAR for the airport indicated that the winds were from 240 degrees at 15knots gusting to 27knots. The visibility was 3 miles in light snow and mist with a few clouds at 1,100 feet, broken clouds at 2,100 feet and overcast clouds at 2,700 feet. The temperature was -1 degree C with a dew point also at -1. PIREPS both before and after the accident reported light to moderate icing from 3,000 to 14,000 feet (NTSB, 2010).

The flight had departed Newark Liberty International Airport at 2118 EST (NTSB, 2010) for the 50-minute flight to Buffalo. The flight had been routine up until that time with the exception of non-standard communication during sterile portions of the flight (below 10,000 feet).

While preparing for the approach at 4000 feet, the first officer asked the captain if the aircraft was accumulating ice to which he responded that there was ice on his side of the windshield. The first officer then responded, “lots of ice” and the captain again commented, “that’s the most I’ve seen – most ice I’ve seen on the leading edges in a long time" (NTSB, 2010). Air traffic control (ATC) continued to monitor the descent of the airplane to 2300 feet (MSL) and at 2212 EST, the flight was cleared for the ILS to runway 23. The crew had the autopilot engaged during this portion of the flight and the airspeed was 180 knots. Approximately three miles from the outer marker, the captain began to slow the airplane toward its final approach speed by reducing engine power towards flight idle. At 2216:21 EST, the first officer lowered the landing gear and selected flaps to 15 degrees as requested by the captain. The airspeed at this time was 145 knots and decreasing. At 2216:27 (six seconds later) the cockpit voice recorder (CVR) recorded a sound similar to the stick shaker and the autopilot disconnect horn that sounded until the end of the recording (NTSB, 2010). The flight data recorder (FDR) that at the time recorded an airspeed of 131 knots. Within .5 seconds of the autopilot disengaging, the FDR showed that the control column moved aft (commanding a pitch up). The power levers were also advanced to about 75% torque (a measure of engine power). The FDR also showed that while the power levers were being advanced the airplane pitched up and rolled to the left approximately 45 degrees and then quickly rolled to the right (NTSB, 2010). Concurrent with the roll, the stick pusher also activated (it would activate two more times) attempting to push the nose of the aircraft down. At 2216:34 the first officer selected the flaps to zero (uncommented by the captain), airspeed at that time was 100 knots (NTSB, 2010). FDR showed that the roll angle reached 105 degrees right wing down before the airplane again began rolling left. The airplane rolled approximately 35 degrees to the left and then began a rapid roll to the right reaching 100 degrees right wing down. At 2216:50, the FDR indicated that the airplane had pitched 25 degrees nose down (NTSB, 2010). Impact with the ground was at 2216:54. From the onset of the stick shaker where the airplane was still flyable to impact with the ground was 26 seconds.

Figure 4. Flight 3407 – One Minute from Impact – Situation Normal *(NTSB, 2010)*


HOT-2: gear's down.
HOT-1: flaps fifteen before landing checklist.
HOT-2: uhhh.

Figure 5. Flight 3407 – 30 Seconds from Impact – Stick Shaker Activation *(NTSB, 2010)*

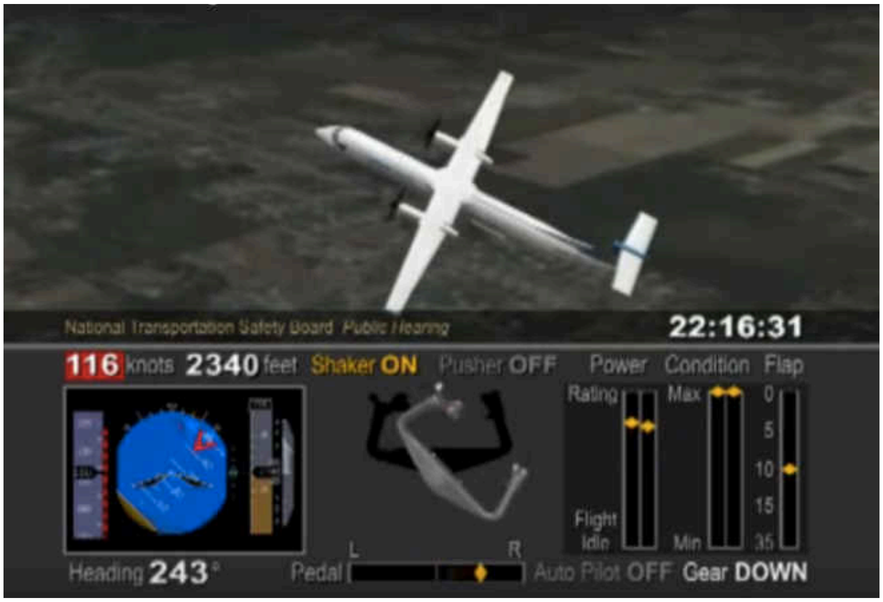


Figure 6. Flight 3407 -27 Seconds from Impact - Roll to the Left (NTSB, 2010)

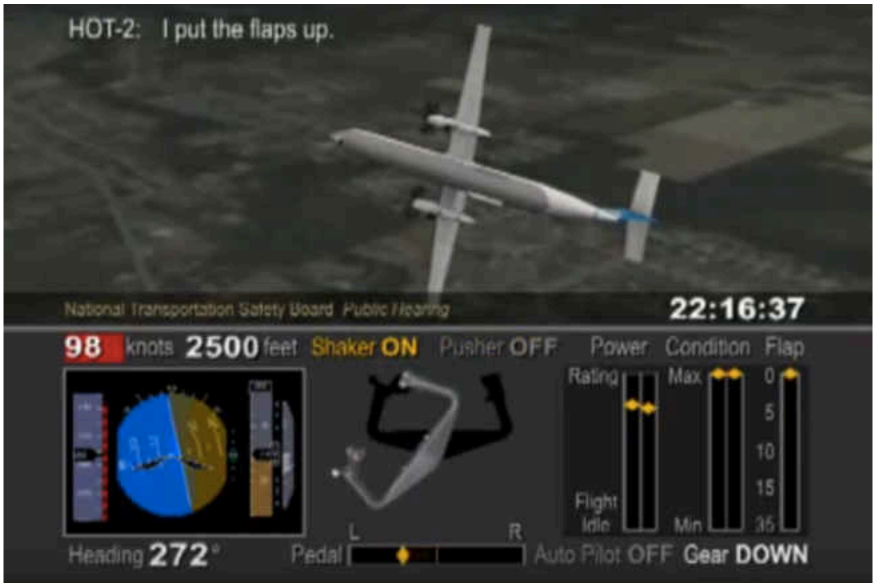

HOT-2: I put the flaps up.

Figure 7. Flight 3407 –21 Seconds from Impact - Roll to the Right *(NTSB, 2010)*

HOT-2: should the gear up?

HOT-1: gear up.

Figure 8. Flight 3407 – 10 Seconds form Impact - Final Roll to the Right and Pitch Down *(NTSB, 2010)*

The accident was investigated by the National Transportation Safety Board (NTSB). An extensive review of both pilot's qualifications was conducted by the Board. The captain had experienced several unsuccessful events during his flying career requiring additional training, however, other pilots who had flown with both the captain and first officer described their performance as "good" (NTSB, 2010).

The NTSB in its findings stated one of the primary causes of the accident was the captain's incorrect actions in response to the stall warning during the approach (NTSB, 2010). It also stated that the icing on the airplane would have resulted in minimal performance degradation (NTSB, 2010). When the stick shaker activated, the captain responded by applying a 37-pound pull force to the control column, which resulted in a nose up elevator deflection. The angle of attack (AOA) increased to 13 degrees with a pitch of 18 degrees. As a result of the power settings and the captain's actions, the airspeed dropped to 125 knots. After the first stick pusher activation, the captain again applied a nose up force on the control column (NTSB, 2010). The captain applied two additional pull forces of increasing magnitude in response to the two other stick pusher activations. The NTSB characterized the captain's actions as "abrupt and inappropriate" (NTSB, 2010).

According to the NTSB, the captain's performance suggest that he was startled by the activation of the stick shaker and responded by making inappropriate control inputs (NTSB, 2010). The NTSB further stated:

> "The captain's failure to make a standard callout or even a declarative statement associated with a recovery attempt and his failure to silence the autopilot disconnect horn (which continued for the remainder of the fight and could have been silenced by pushing a button on the control wheel) further suggest that he was not responding to the situation using a well-learned habit pattern. The first officer was not providing guidance consistent with an understanding of the situation (NTSB, 2010)".

A scientist at the NASA-Ames Research for Aerospace Human Factors stated that people under stress might not respond appropriately to events in their environment (NTSB, 2010). The captain's response to the stick shaker should not have required cognitive effort to make the correct inputs or callouts (NTSB, 2010). In a possible explanation to the captain's response, the NTSB cited Colgan's training on icing in which a video on tail plane stalls was shown (NTSB, 2010). The recovery that the captain attempted was similar to that which should be taken during a tail plane stall, however, the aircraft itself presented no evidence of such an event (NTSB, 2010). It is more probable that in reaction to a startle event, the captain chose the incorrect cognitive pathway for resolution and was never able to correctly diagnose the true issue with the airplane.

Aircraft Accident Air France Flight 447

Air France flight 447 (AF447) also represents a case where startle of the crew may have adversusly affected initial decision making resulting in a loss of the aircraft (BAE, 2012). AF447 was a regularly scheduled flight from Rio de Janeiro to Paris. AF447, an A330, departed Rio on June 1st, 2009 carrying 216 passengers, three pilots, and nine flight attendants. Routing of the flight was over the central Atlantic Ocean. The flight proceeded normally for the first two hours and was flying level at 35,000 feet.

Approximately 2:10:05, the aircraft encountered freezing precipitation which obstructed the pitot probes (BAE, 2012). The loss of the pitot probes affected the autoflight system and the cockpit airspeed indications. The total time from onset of the pitot issue to impacting the ocean was 4:23. When the autoflight system disconnected, the pilot flying (PF) began applying a nose up command on the sidestick. The cockpit speed indications dropped from 275 knots to 60 knots (typical of an icing event). At 2:10:16 the pilot not flying (PNF) stated “we’ve lost the speeds” then “alternate law protections” (BAE, 2012). The PF made rapid and high amplitude roll control inputs (from stop to stop). He also made an additional nose-up input that increased the airplane’s pitch attitude up to 11 degrees (BAE, 2012). The airplane was in a climb through 37,700 feet at this point. At 2:10:36 the airspeed on the left side of the cockpit became valid as the ice melted in the pitot probe. Airspeed at this point was 223 knots which represented a loss of 50 knots. The PF reduced the pitch of the airplane momentarily at 2:10:47 however, he then resumed a pitch-up beyond 10 degrees and the airplane again began to climb. This pitch caused the airplanes stall-warning system to trigger in a continuous manner (BAE, 2012).

The PF selected maximum thrust on the engines and made additional pitch up inputs towards 13 degrees. Approximately 15 seconds later the right side airspeed indicator became valid and recorded an airspeed of 185 knots (BAE, 2012). The PF continued to command a pitch up and the airplane reached a maximum altitude of 38,800 feet and an angle of attack (AOA) of 16 degrees. At 2:11:42 the captain re-entered the cockpit from a rest break. At that time all three airspeed indicators were displaying valid airspeed data. Also around this time, the airplane was descending through 35,000 feet with an AOA of 40 degrees, which resulted in a vertical speed of -10,000 feet per minute (fpm). The airplane was also experiencing roll oscillations exceeding 40 degrees (BAE, 2012).

Due to the extremely low airspeed, the stall warning ceased. The PF momentarily reduced the pitch of the airplane which again triggered the stall warning as the plane gained a little speed. Unfortunately, the PF resumed the commanded pitch up and the AOA approached 35 degrees. The last recorded data for the airplane was at 2:14:28. Data indicates a vertical speed of -10,912 fpm with an airspeed of 107 knots.

Figure 9 is a graphical reproduction of the flight parameters from the flight data recorder. The timeline for this diagram is the first 50 seconds of the event where the aircraft goes from controlled flight to a descent rate of over 10,000 feet per minute. The chart shows the aggressive handling by the first officer in both pitch and roll. It also shows when the airspeed information became valid.

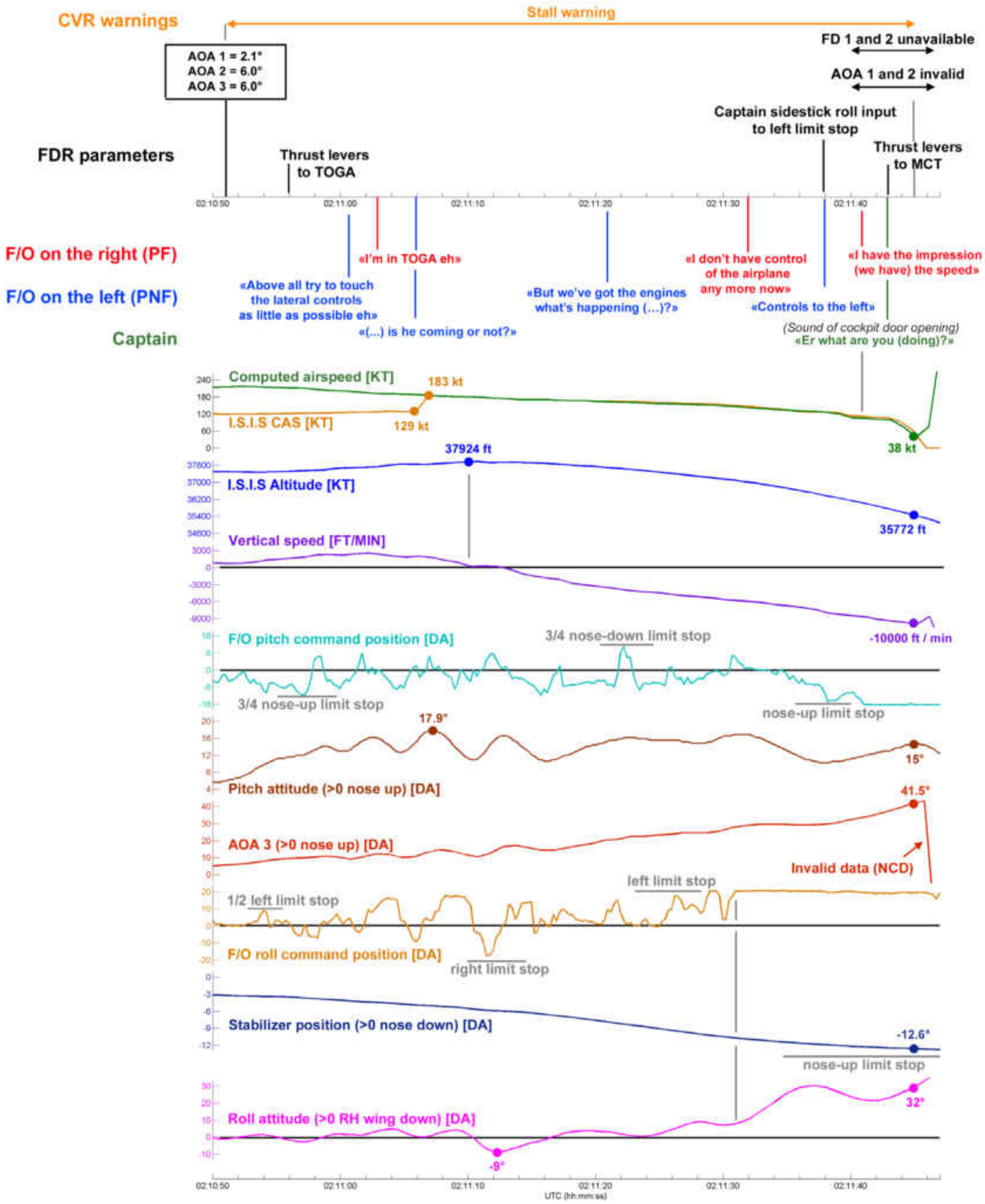


Figure 9. Parameters from 2:10:50 to 2:11:46 *(BAE, 2012)*

Figure 10 shows that the airspeed indications were normal just prior to the event. It also shows that the airspeed indications (for the first officer) may have not been reliable for a period of 40 seconds. It also shows that once the airspeed indications were valid, a constant decrease in speed occurred until the end of the event.

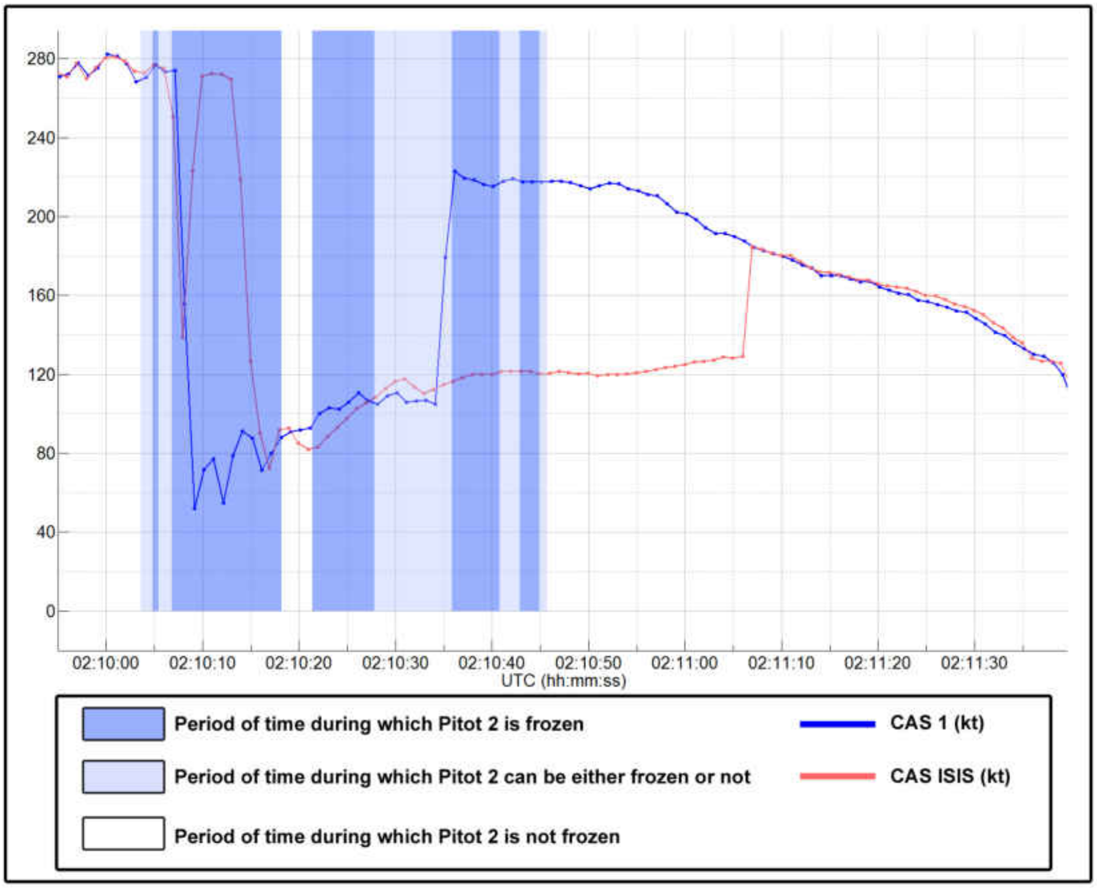


Figure 10. Evolution of Airspeed and Pitot Icing *(BAE, 2012)*

Figure 11 summarizes the event from the onset of the frozen pitot tubes until impact with the ocean. Large variations in pitch and angle of attack can be seen throughout most of the event. The aircraft is completely stalled for the last minute of the event, finally impacting the ocean is a nose high with almost no forward airspeed and at a rate of descent of -15,000 feet per minute.

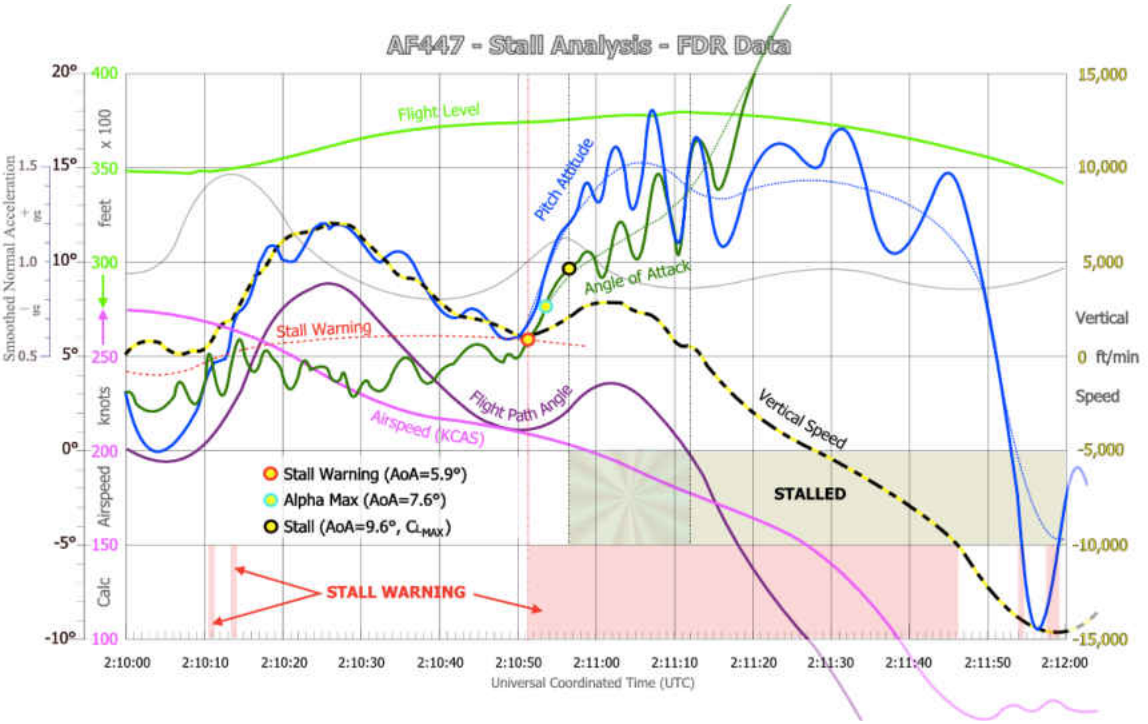


Figure 11. AF447 FDR Data *(MM43, 2011)*

In its findings, the BAE studied 13 incidents related to icing and unreliable airspeed. Air France had four such cases in their history with the A330 aircraft. The BAE determined that in less than one minute after the autopilot disconnected, the airplane exited its flight envelope following inappropriate pilot inputs (BAE, 2012). The airplane went into a sustained stall as signaled by the stall warning system and strong airframe buffet. Even though the stall warning sounded for 54 seconds, neither pilot made any reference to the stall warning or the associated buffet. The crew never applied the stall recovery maneuver. The incident startled the crew and they had difficulties handling the airplane (BAE, 2012). The excessive pitch and vertical speed added to the erroneous indication and emergency, caution, and monitoring (ECAM) messages, which added complexity in the diagnosis of the situation. The crew likely never understood that it was a simple loss of airspeed data (BAE, 2012).

Conclusion

Although not intended to be all encompassing, this literature review seeks to provide the reader a broad background on which to base this study. The review discussed startle effect and the current understanding of its effect upon decision making. Emergencies, where flight crews made the incorrect initial decision, become progressively more difficult to successfully manage as pilots often selectively filter information to confirm their initial decision. Decisions evolve from past experiences and knowledge where bits of similar experiences are pasted together in the model generation phase.

Decision-making is a complex process that researchers are only beginning to understand. It is known that when pilots are startled by a sudden emergency that their decision-making and subsequent performance can be adversusly affected. The startle effect may lead to a breakdown in crew coordination and puts additional cognitive load on the individual pilot. With over-reliance on automation, crews may not be well equipped to handle a sudden inflight emergency that requires the use of hand flying skills. The startle effect may, in some cases, result in incorrect model generation/selection with incorrect decisions and/or actions being applied to an emergency situation. Boeing (2012) suggests that inflight loss of control is the single highest category for airline fatalities. Startle effect, cognitive overload, and crew breakdown can all be seen in both the Air France and Colgan accidents. In many cases, simulator training in conjunction with deliberate practice has been demonstrated to increase crew performance. Deliberate practice has been widely applied as an industry solution to other inflight emergencies and applying it startle effect may be an effective way to mitigate some of its inherent risks.

CHAPTER II

METHODOLOGY

Introduction

This study was a mixed methodology study focusing on whether startle training could help successfully mitigate the cognitive gap that exists during a startle event. The first part of the study was a survey given to each of the participating pilots. The second part of the study evaluated the crews as they flew one of the selected maneuver sets and was quantitative in nature. The simulator part of the study was a quasi-experimental design with crews that did not receive startle training serving as the control group. Each participating crew was evaluated either a low or high altitude scenario depending on the day of the week. Randomly selected crews received training on handling the aircraft during a startle event. The training consisted of both a briefing and simulator practice. Practice in the simulator was equal for the crews in the trained group lasting approximately one hour. The briefing consisted of personal instruction using a power point presentation discussing the proper pitch, power, bank, and time recognition (see Appendix B). The briefing ended with a mnemonic device that pilots were expected to use and verbalize both in the practice events and the evaluation event. Simulator practice consisted of a startle event not related to the evaluation profiles. It was intended to have equally trained and untrained crews to compare performance between the groups.

The first test scenario was a low altitude and low fuel profile. The scenario degrades with a system failure that causes a missed approach and a resequence for landing. Time pressure, low fuel, and the unexpected missed approach combine to form the startle event and event evaluation begins at the missed approach.

The crew flew a standard arrival procedure into Newark Liberty International Airport (KEWR). The routing for the arrival had both lateral and vertical restrictions and would be considered a routine procedure for approaching the airport. The specific arrival chosen for this study was the DYLIN arrival (see Appendix D). The weather at KEWR combined with traffic saturation has caused holding (at the METRO intersection). This holding is unexpected by the crew resulting in somewhat of a low fuel situation that adds an initial stress element to the scenario. The crew is cleared out of holding for the instrument approach to runway 4R (see Appendix D) with approximately one hour of fuel remaining. The exact amount of fuel depends on the aircraft in the scenario (See Appendix D). The vectors and the initial part of the instrument approach to runway 4R were normal. The weather was instrument flight rules (IFR) with a 500-foot overcast ceiling and a visibility of one mile.

When the crew selected the landing gear down, around approximately 2000 feet on the approach, one of the landing gear fails to extend. It was expected that the crew will execute a missed approach at this point in order to try and rectify the landing gear issue. The landing gear malfunction is the startle event and time pressure due to low fuel serve to add stress to the crew.

The second test scenario was a high altitude profile. The crew was briefed that they are on a flight that terminates in KEWR, with routing via the DYLIN arrival. The flight is at 35,000 feet (FL350), with a descent planned via the arrival. There have also been reports of light icing descending into KEWR. Pilots are in instrument conditions with light turbulence. The scenario involves loss of the aircraft's air data system, which disables many of the auto-flight systems. With the air data loss, an engine fire warning was introduced. The air data loss was the startle event and the engine fire warning added a distractor and stress to the crew. Evaluation began at the loss of air data. The air data loss renders the aircraft's speed, altitude, and vertical trend unreliable with the side effect of the autoflight system automatically disconnecting. The air data interruption was of short duration and only a few seconds elapsed before instrument indications return to normal. The autoflight disconnection forces the crew into a hand flying situation and the engine fire warning serves as both a distraction and a startle event.

Data collection consisted of crew performance as it relates to aircraft control for each scenario. Each scenario is made up of five sub-tasks which were evaluated and used to determine an overall score. The scoring methodology was taken directly from participating airlines' FAA approved advanced qualification program (AQP) evaluation manual. Each pilot group was compared using a one-way ANOVA against the FAA proficiency standard and then compared to the other group for significance. T-tests were performed on the different maneuver sets to determine if the proximity of the maneuver affects the pilot's ability to successfully fly the aircraft. If the study hypothesis is correct, the untrained crews should show a significant statistical difference as compared to the standard pilot performance as defined by the FAA. In addition, regression analysis was performed on various aspects of the collected data to gain an insight as to where the variability lies.

Subjects

The population for this study were professional pilots of an FAR 121 commercial air carrier. Furthermore, the study focused on pilots of scheduled passenger airlines. Flight crews from the participating airline were asked during their recurrent training cycle if they wished to participate in this study. They were selected based on their willingness to volunteer for this study. Selection of crews generally occurred during their final day of training when there was often extra simulator time available. If the crew volunteered, they were requested to fill out a survey on their experience and perceptions of their flying abilities, especially during unusual events. The crews were informed that there was no personal data kept and no data linking an individual to their performance.

Beta Testing

A small group test was completed on approximately 15 subjects. The nature of the beta test was to determine an estimated effect size of the independent variable. The effect size was used to complete a power analysis. In addition, the beta test group was used to establish the simulator parameters for items such as weight, fuel, and induced problems. The beta test results are reported here and not in the findings section as they are not considered part of the actual test data. IBM SPSS Statistics (SPSS) was used to analyze the overall training effect and then for each scenario (low and high). Tables of the initial results are listed below. Table 1 lists the basic descriptives. A total of 15 crews participated in the beta test.

Table 1. Descriptive Statistics Beta Group

Dependent Variable: Overall score collapsed across low and high altitude

Crew training provided	High or low altitude scenario	Mean	Std. Deviation	N
Yes	High Altitude	4.0000	.00000	5
	Low Altitude	4.5000	.57735	4
	Total	4.2222	.44096	9
No	High Altitude	2.3333	.57735	3
	Low Altitude	2.6667	.57735	3
	Total	2.5000	.54772	6
Total	High Altitude	3.3750	.91613	8
	Low Altitude	3.7143	1.11270	7
	Total	3.5333	.99043	15

The beta test found that the training effect was significant with p = .00 when the low and high altitude groups (trained and untrained) where compared. The results are listed in table 2.

Table 2. Tests of Between-Subjects Effects Beta Group

Dependent Variable: Overall score collapsed across low and high altitude

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	11.400[a]	3	3.800	17.914	.000	.830	53.743	1.000
Intercept	163.209	1	163.209	769.414	.000	.986	769.414	1.000
Training	10.970	1	10.970	51.716	.000	.825	51.716	1.000
HighLow	.622	1	.622	2.932	.115	.210	2.932	.346
Training HighLow	.025	1	.025	.117	.738	.011	.117	.061
Error	2.333	11	.212					
Total	201.000	15						
Corrected Total	13.733	14						

a. R Squared = .830 (Adjusted R Squared = .784)

b. Computed using alpha =.05

Table 3 below compares the trained and untrained groups for only the low altitude scenario. The descriptive statics and ANOVA results are in Tables 3 and 4.

Table 3. Descriptive Statistics Low Altitude Scenario Beta Group

Dependent Variable: Low Altitude Scenario

Crew training provided	Mean	Std. Deviation	N
Yes	4.5000	.57735	4
No	2.6667	.57735	3
Total	3.7143	1.11270	7

Table 4. Tests of Between-Subjects Effects Low Altitude Scenario Low Altitude Beta Group

Dependent Variable: Low Altitude Scenario

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	5.762[a]	1	5.762	17.286	.009	.776	17.286	.908
Intercept	88.048	1	88.048	264.143	.000	.981	264.143	1.000
Training	5.762	1	5.762	17.286	.009	.776	17.286	.908
Error	1.667	5	.333					
Total	104.000	7						
Corrected Total	7.429	6						

a. R Squared = .776 (Adjusted R Squared = .731)

b. Computed using alpha =.05

The same comparison was done on the beta group for the high altitude scenario. The results are listed in Tables 5 and 6. Like the low altitude beta group, the main effect of training showed significance.

Table 5. Descriptive Statistics High Altitude - Beta Group

Dependent Variable: High Altitude Scenario

Crew training provided	Mean	Std. Deviation	N
Yes	4.0000	.00000	5
No	2.3333	.57735	3
Total	3.3750	.91613	8

Table 6. Tests of Between-Subjects Effects High Altitude - Beta Group

Dependent Variable: High Altitude Scenario

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power^b
Corrected Model	5.208^a	1	5.208	46.875	.000	.887	46.875	1.000
Intercept	75.208	1	75.208	676.875	.000	.991	676.875	1.000
Training	5.208	1	5.208	46.875	.000	.887	46.875	1.000
Error	.667	6	.111					
Total	97.000	8						
Corrected Total	5.875	7						

a. R Squared = .887 (Adjusted R Squared = .868)

b. Computed using alpha =.05

Analysis suggest that between 73% and 86% of the variability can be accounted for due to the training effect. G-Power was then used to calculate a Cohen's D to estimate the training effect size. The calculated effect size was .33 which reflects a medium effect size. The given effect size was then used to estimate the number of subjects needs to ensure an adequate sample size which with the calculated effect size is between 17 and 60 crews.

Table 7. Cohen's D Calculation Beta Group

Scenario	Low Altitude	High Altitude
Mean	3.167	3.375
Std. Deviation	0.916	1.112
Sample Size	7.0	8.0
Result	Cohen's d = (3.714 - 3.375) ⁄ 1.018725 = 0.332769.	

Cohen's d = (M2 - M1) ⁄ SD_{pooled} SDpooled = √((SD12 + SD22) ⁄ 2)

Sample Groups

The plan was to evaluate 30 - 60 crews. This allowed for enough data to be collected even if the effect size in the rating scale is small to medium (Cohen's d = .33 / r = .19). The number of pilots were chosen in order to gain a statistically significant sample approximating the skill level of the general professional pilot population. There were four main groups for comparison. The groups were defined as follows:

1. Low Altitude – No Training (LANT)
2. Low Altitude – Training (LAT)
3. High Altitude – No Training (HANT)
4. High Altitude – Training (HAT)

Equipment

This study used professional airline pilots flying two different scenarios in an FAA approved Level-D full flight simulator (FFS). Simulators that were utilized include A320, B737, B757, B767, B777, B787, and B747. The scenarios were flown by a crew consisting of a captain and first officer, similar to what would happen in actual line operations. The training that some of the crews received was broad-based and not aircraft specific.

Data Collection Methods/Procedures

Data collection for this study focused on two parts: a survey and observed simulator performance data. The survey consisted of approximately 10 multiple-choice questions that were evaluated using a Lickert scale. Questions focus on such items as outside flying, previous military and/or aerobatic flying, and general hand flying attitudes.

The simulator portion of the study involved flying maneuver profiles in an FAA approved Level-D full flight simulator. The maneuver sets were evaluated by the principle investigator. The investigator is a former instructor pilot on multiple transport category jet aircraft with over 11 years of professional instructional experience. The maneuver evaluation criteria were developed in accordance with airline training procedures and protocols as set forth by the FAA. The specific evaluation criteria were adopted from an FAR 121 passenger Advanced Qualification Program (AQP) program (with permission), which was approved by the FAA. The criteria match closely with evaluation standards set forth by the FAA in the Airline Transport Pilot (ATP) practical test standards (PTS). The crews were evaluated on the success of the maneuver as described in Tables 8-10.

Table 8. High Altitude Analysis

Score	Criteria
5	Crew performance was excellent in both aircraft handling and problem diagnosis. The crew had minimal altitude and heading changes, recognized the issue promptly and applied the correct mitigation strategies.
4	Crew performance was good. Problem was correctly diagnosed with pitch and roll not exceeding 5 deg/50 feet.
3	Crew performance was average. Some difficulty diagnosing the problem. Pitch and roll not exceeding 10 deg/ 100 feet.
2	Crew performance was below average. Problems and/or confusion diagnosing the problem or misdiagnosed of the problem. Major deviations in pitch and roll more than 20 deg/200 feet.
1	Crew performance was unacceptable. The crew could not diagnose the problem and misdiagnosed the problem. Excessive deviations and handling of the aircraft.

Table 9. Low Altitude Analysis

Score	Criteria
5	The crew remained well within standards and performance was exemplary. The crew recognized the issue and handled promptly while recognizing the deteriorating fuel state of the aircraft.
4	The situation was well handled with the safety of the flight not in jepordy. The crew was aware of the time pressures and the fuel state and mitigated both.
3	The flight landed safely with no major deviations for SOPs with at least 30 minutes of fuel.
2	Landed the aircraft in less than desirable conditions with regards to configuration, fuel and time management.
1	The pilot committed major deviations from standards that were not promptly corrected and/or were unsafe, or was unable to perform the maneuver/task without assistance. The pilot crashed or lost control of the aircraft.

Table 10. Evaluated Factors and Seat Positions

	High Altitude	Low Altitude
Factors	Problem Diagnosis	Missed Approach
	Pitch	Irregular Checklists
	Roll	Time Management
	Altitude Control	Fuel Management
	Overall Control	Approach and Landing
	Overall Score	Overall Score
Seat Position	Pilot Flying	Pilot Flying
	Pilot Monitoring	Pilot Monitoring

Data Analysis and Statistical Modeling

Analysis was planned for within groups and between groups comparisons with further analysis of significant factors. An additional analysis compared the training groups and non-training groups versus the FAA standard. Overall factor and seat position scores were also compared to the FAA standard, which for this study was a grade of three as described above in the data collection section. The statistical modeling program SPSS was primarily used to analyze the data. In addition, the survey data was compared and analyzed to see if any significant correlations can be determined. ANOVA, linear regression, and post-t- tests were the primary statistical models used.

Within Group Comparison

The first data set for analysis were the within group comparisons. Analysis was conducted between Low Altitude Trained Group (LAT) and Low Altitude Non-Trained Group (LANT) and the High Altitude Trained Group (HAT) and the High Altitude Non-Trained Group (HANT). The first part of the analysis used the overall grade score (see Appendix 1). Analysis looked for significant findings within each group using a one-way ANOVA with the alpha level set at .05. Comparison between the overall grade and the FAA standard grade (3) was also compared. A second round of analysis occurred for the contributing factors of the overall grade. Linear regression was used to determine what part of the variance each factor (if any) are significant. Again the alpha level was .05. Finally, as part of the within groups comparison, the seat position (captain or first officer) was tested for significance as a contributor to the overall grade using regression.

Between Group Comparison

The between-group comparison was similar to the within group comparison using the same tests and tools. The between groups was between LAT and HAT followed by LANT and HANT groups. The analysis sought to determine significant findings of the final grade, using a one-way ANOVA with an alpha level of .05. As with the within group comparison, linear regression was used to analyze both the contributing factors and the seat position. Comparison to the FAA standard was conducted in the between group comparison.

Comparison versus a Known Standard

The final set of comparisons was conducted by collapsing across groups (trained versus untrained, and then comparing against a known set standard. The set standard was determined by the tolerances set forth by the FAA for an Airline Transport Pilot (ATP) certificate. The data collection section above describes the grading standards as set by the FAA and airline policies, which since they are approved are also part of the FAA standard. Crews receiving, at least, an overall grade of (3) were considered to meet the FAA standard. Any grade below (3) was considered below standard.

The training groups (LAT and HAT) and non-training groups (LANT and HANT) were collapsed and then compared to the FAA standard using a one-way ANOVA with an alpha level of .05. Comparison of the collapsed groups was also compared to the survey responses using the Pearson Correlation test.

Protection of Human Subjects

The crews in this study were exposed to unusual but not extraordinary aircraft failures in the simulator. These failures are regularly practiced during initial and recurrent training. The crews are accustomed to having their performance evaluated. There is also a minimal risk of performing in front of a colleague if the performance is substandard even though the evaluation will be as a whole crew. This is mitigated by not identifying a particular pilot with an individual performance. The volunteer subjects were also encouraged not to evaluate each other’s capability based on this testing scenario. Finally, crews will be requested to refrain from talking about the testing with other crews as to not compromise the test data. There will be no identifiable link between the performance of the crew and any individual crew member (a condition set by the participating airline in granting use of their simulators). Individual performance will not be reported to any airline or entity outside the research project.

Since minimal risks will be involved in the study, the subjects will be informed that the study will include various maneuvers that will be flown in the simulator and that the crew's performance will be analyzed. There will be a check box on the survey form indicating that the participants have volunteered for the study. The specific language is as follows: " By checking this box, I agree that I have volunteered for this study and have felt no undue pressure from the airline, the University of North Dakota, or the principle investigator to participate. I have also been informed that no data will be kept linking any simulator performance to a specific pilot. Data collected is for this research project only and will not be reported to any entity or airline. Final aggregate results may be viewed in the published dissertation that will be available at the University of North Dakota Chester Fritz Library. The researcher has informed me that I will fly as part of a crew and may encounter some unusual situations in the simulator. I also understand that I have the right to refuse participation or withdraw from the study at any point without a change in relationship with my airline, the University of North Dakota or the research team." All participants acknowledged and signed the participation consent form.

CHAPTER IV

RESULTS

This study consisted of two main parts, a survey and a flight evaluation. There was data recorded in an FAA approved Level-D flight simulator, flown by pilots for a major US based passenger airline. Volunteer crews were asked to fly one of two different scenario profiles. Random crews received training that consisted of a briefing and simulator practice. The training sought to mitigate the negative cognitive effects following a startle event. The data mainly focuses on the effect the training had on the trained pilot group. Analysis consisted of both within and between main groups with regression analysis on the contributing factors that made up the maneuver set scores. Crews were presented with either a low altitude and low fuel scenario or a high attitude scenario with a loss of air data. The survey was conducted in order to gain a perspective into how pilots at major airlines fly their aircraft, and how they perceive their own flying skills.

Demographics

Forty crews who flew for a U.S. Global passenger airline participated in the study. All of the subjects were active line pilots and volunteers. The pilots flew as a crew consisting of a Captain and First Officer and had flown in their respective aircraft for at least one year. Crews were also divided by which profile they flew and whether they received training and practice prior to flying the profile scenario. Each scenario (low or high) was flown by 20 crews. In addition, crews were separated by what type of aircraft that they flew. There were 21 wide-body aircraft crews (B747, B787, B777, B767) and 19 narrow-body aircraft crews (B737, A320, B757).

Survey Responses

The survey was divided into two distinct parts: a pilot's experience, and their perception of their own skills. The pilot flying (PF) the scenario was asked to complete the survey. The first survey question asked if the pilot flew outside of their current job in another professional manner such as flight instructing. The results are displayed in Table 11 and Figure 12

Table 11. Flying Outside of Professional Job

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	6	15.0	15.0	15.0
	No	34	85.0	85.0	100.0
	Total	40	100.0	100.0	

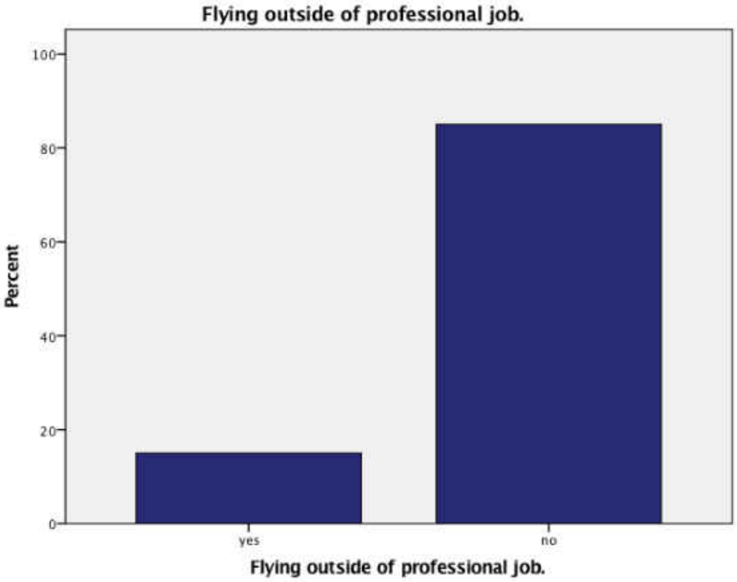


Figure 12. Outside Flying

Survey responses indicated that only 15% of the pilots flew outside of the current job. Outside flying generally consists of less sophisticated aircraft that require more routine flying skill practice. Since only 15% indicated that they flew outside of their airline job, the results were not considered significant.

The next survey question asked whether the pilot flew in the military. Of the survey responses, 32.5% indicated that they have flown in the United States military. The results are displayed in Table 12 and Figure 13

Table 12. . Did You Fly in the Military?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	13	32.5	32.5	32.5
	No	27	67.5	67.5	100.0
	Total	40	100.0	100.0	

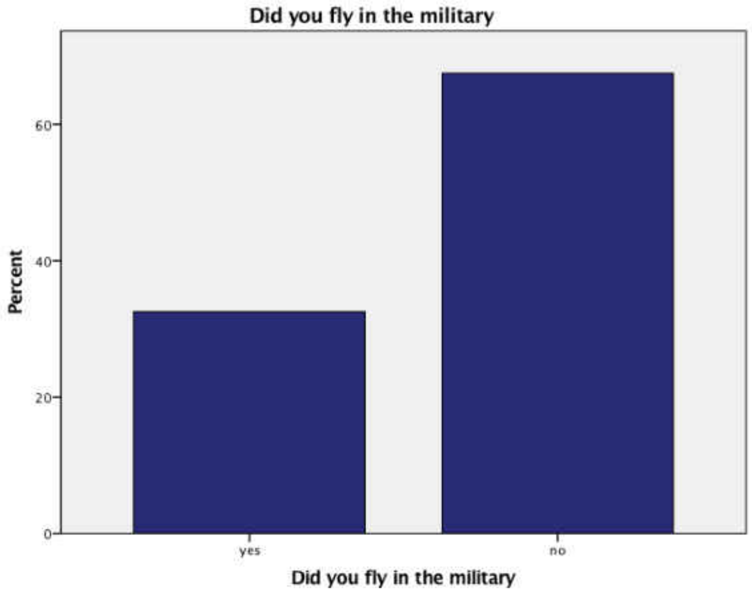


Figure 13. Civilian versus Military Flying

Anecdotal personal interviews with military pilots suggested that those exposed to military flying have a greater exposure to startle events and in some cases have developed coping mechanisms.

The next survey question asked if the pilot had any type of formal aerobatic training. This type of training may indicate better recognition of unusual attitudes and lead to a more effective response to an unusual event. The results are displayed in Table 13 and Figure 14.

Table 13. Do You Have any Formal Aerobatic Training

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	23	57.5	57.5	57.5
	No	17	42.5	42.5	100.0
	Total	40	100.0	100.0	

More than 57% of the pilots indicated that they had received some type of formal aerobatic training.

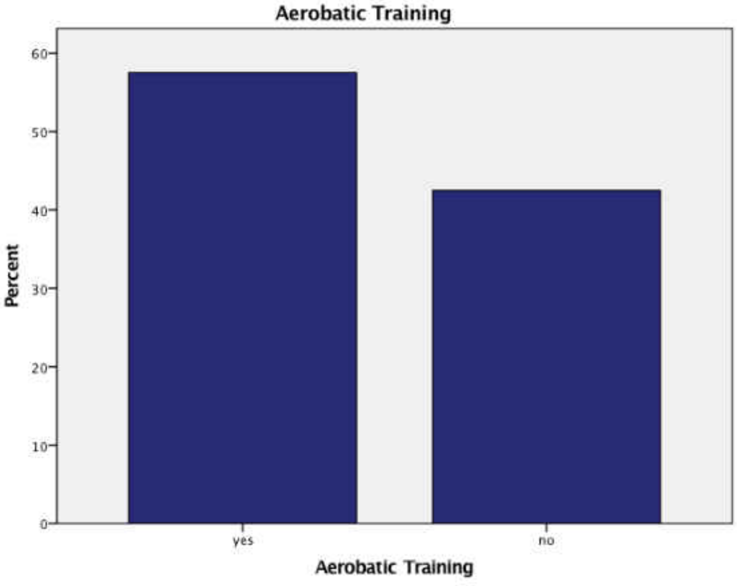


Figure 14. Aerobatic Training

The final survey question related to pilot experience, asked whether the pilots had ever encountered an unusual event that they would describe as “startling”. This question was asked to gage how many pilots have experienced events (while flying) that caught them by surprise. The results are displayed in Table 14 and Figure 15.

Table 14. Startle Events

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	32	80.0	80.0	80.0
	No	8	20.0	20.0	100.0
	Total	40	100.0	100.0	

A majority of the pilots (80%) indicated that they had been startled while flying, leading to the conclusion that startle is somewhat common among professional fight crews.

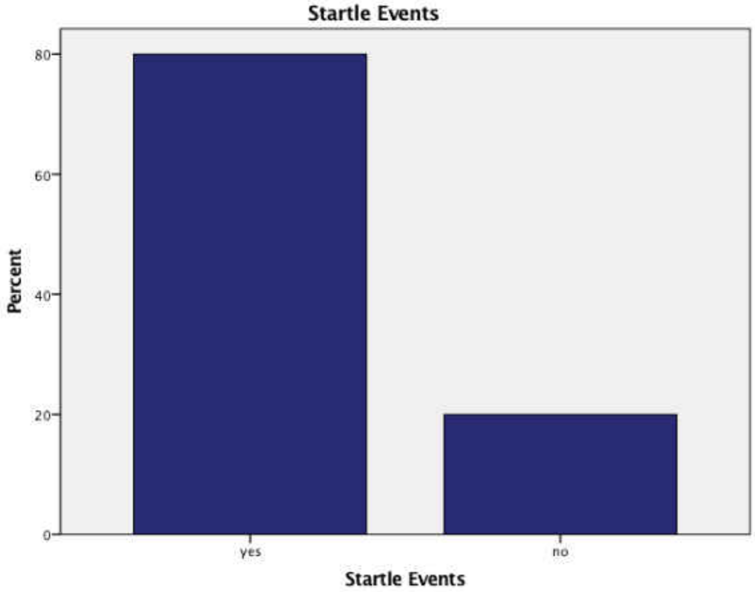


Figure 15. Startling Events

The next section of the survey sought to gain a perspective on how pilots generally flew the aircraft in normal line operations. This section of the survey asked the pilots to rate the questions based on a sliding scale of agreement (1-5) from "strongly agree" to "strongly disagree". The pilots were asked to select their agreement with the survey questions. There was no option to select responses outside of the five standard ones.

The first statement asked whether a pilot knows the proper pitch and power settings for phases of flight such as cruise and approach. This is important because if various flight instruments are lost, safe flight can be continued with just a pitch and power setting. No pilots disagreed with this statement with 50% strongly agreeing that they knew the correct pitch and power settings. The results are described in Table 15 and Figure 16.

Table 15. I Know the Proper Pitch and Power Settings

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Agree	20	50.0	50.0	50.0
	Somewhat Agree	15	37.5	37.5	87.5
	Neutral	5	12.5	12.5	100.0
	Total	40	100.0	100.0	

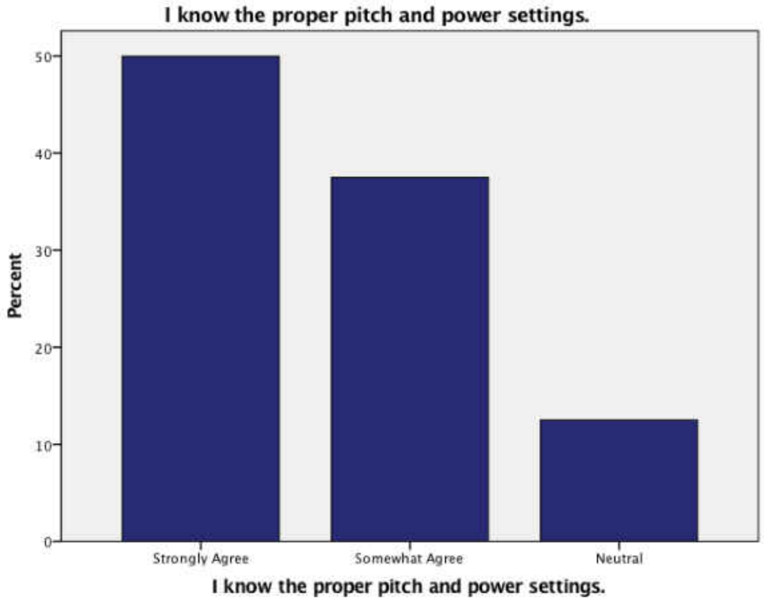


Figure 16. Pitch and Power Settings

When combined with “somewhat agree”, positive responses provided by the pilots recorded at 87% . None of the pilots indicated that they disagreed with the statement. Responses indicate that a majority of pilots believe that they know the correct pitch and power settings for the phase of flight.

The next survey question asked if the pilots often hand-flew the aircraft during departure and approach below 10,000 feet. These phases of flight often contain the most complex aircraft maneuvering. Changes in altitude, speed, and course are routine in these phases. Departures involved changes in routing while climbing and configuring the aircraft for high speed flight. Arrivals involve a similar sequence only in reversus order. The results are displayed in Table 16 and Figure 17.

Table 16. Hand Flying Below 10,000 Feet

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Agree	27	67.5	67.5	67.5
	Somewhat Agree	9	22.5	22.5	90.0
	Somewhat Disagree	4	10.0	10.0	100.0
	Total	40	100.0	100.0	

Responses to this question were indicate that 67.5% strongly agreed with this statement and 22.5% somewhat agreed. Only 10% of the pilots disagreed with the statement. There was no neutral or strongly disagree statements. This indicates that most pilots are hand flying the aircraft below 10,000 feet.

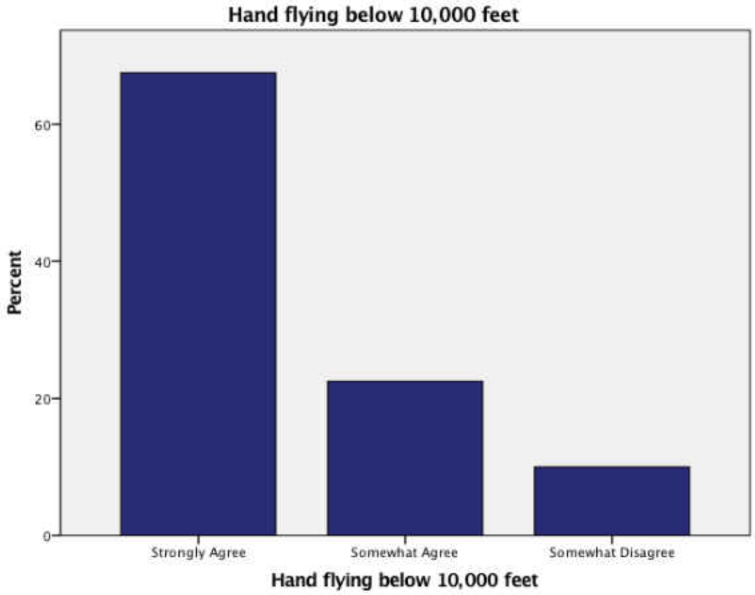


Figure 17. Hand Flying Below 10,000 Feet

Mental rehearsal of different flight scenarios has been found to helpful in shaping pilot responses to unusual situations. At the major airline studied for this research, pilots are required to view and participate in organized "chair flying" during the recurrent training cycle. The next survey question asked whether the pilots extended this practice outside of their recurrent training cycle. The results are displayed in Table 17 and Figure 18.

Table 17. Chair Fly Scenarios to Help Determine Courses of Action

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Agree	6	15.0	15.0	15.0
	Somewhat Agree	17	42.5	42.5	57.5
	Neutral	11	27.5	27.5	85.0
	Somewhat Disagree	3	7.5	7.5	92.5
	Strongly Disagree	3	7.5	7.5	100.0
	Total	40	100.0	100.0	

The pilots indicated that they somewhat agreed to this statement 42% of the time. The next largest group was neutral representing 27.5% of the responders.

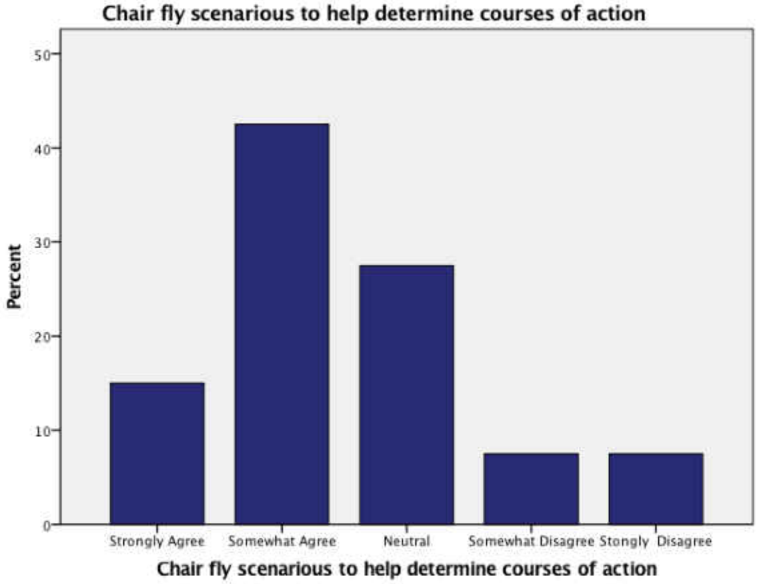


Figure 18. Chair Flying

The statement recorded responses in all categories. Responses indicated that a 57.5% of the pilots use this practice.

Being able to fly the airplane without advanced automation has been shown as a key element in recovering from an unusual situation. When an aircraft is upset (outside of the normal flight envelope), the automation will often disconnect (United Airlines, 2016). This survey question asked if the pilots were comfortable flying the aircraft without the use of the flight director, autothrottles, and map mode. The results are displayed in Table 18 and Figure 19.

Table 18. Comfort Flying Raw Data

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Agree	17	42.5	42.5	42.5
	Somewhat Agree	15	37.5	37.5	80.0
	Neutral	3	7.5	7.5	87.5
	Somewhat Disagree	5	12.5	12.5	100.0
	Total	40	100.0	100.0	

There were no pilots who strongly disagreed with this statement. Pilots selecting strongly agree and somewhat agree were 80% of the responses.

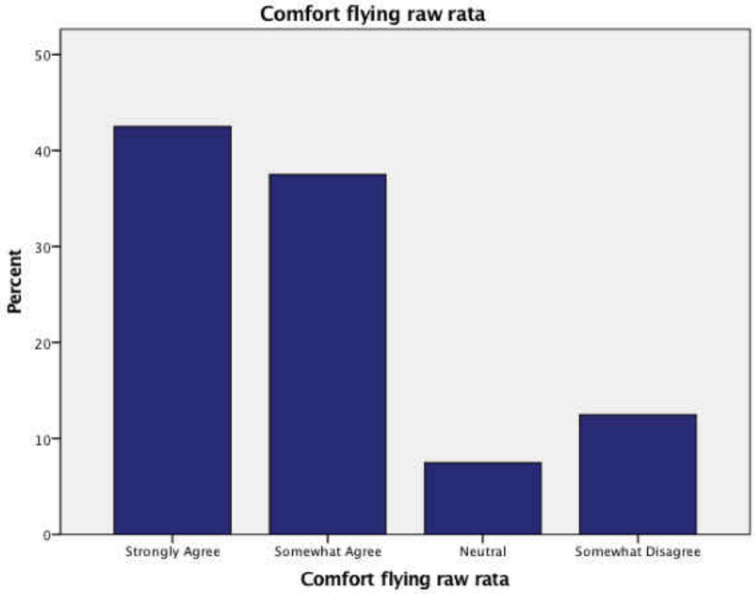


Figure 19. Raw Data Flying

These responses indicate that most pilots feel that they are comfortable flying the aircraft with raw data only.

Hand flying the airplane during the day in good weather is common, however deliberate practice in conditions other than day visual flight rules (VFR) is important in maintaining flying skills. The next survey questions asked pilots if they hand flew in various conditions. The results are displayed in Table 19 and Figure 20.

Table 19. Often Practice Raw Data Skills

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Agree	22	55.0	55.0	55.0
	Somewhat Agree	10	25.0	25.0	80.0
	Neutral	7	17.5	17.5	97.5
	Somewhat Disagree	1	2.5	2.5	100.0
	Total	40	100.0	100.0	

Of the pilots surveyed, 80% either strongly agreed or somewhat agreed with the statement. A larger number of pilots indicated that they were neutral 17.5% with this statement when compared to the other questions.

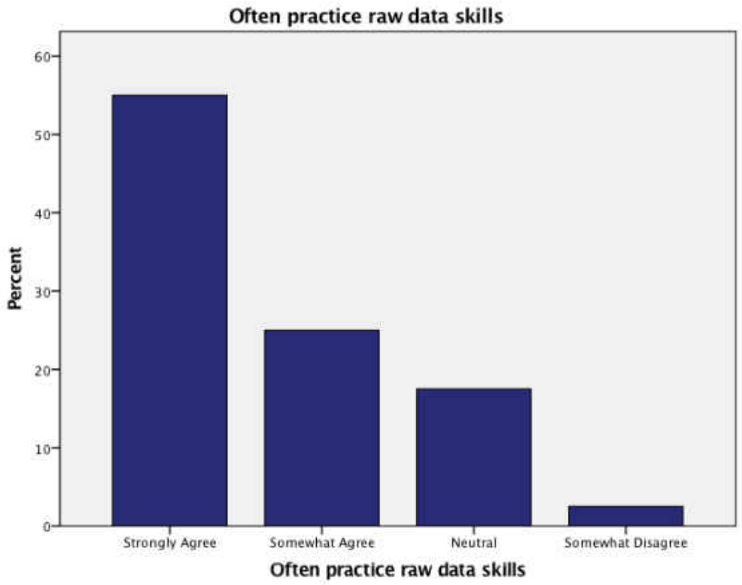


Figure 20. Skills Practice

The responses indicate that most of the pilots practice hand flying under various conditions.

The final question asked if the pilots used the autopilot for a majority of the flight above 1000 feet. This question is in contrast to the 10,000-foot altitude hand flying question and sought to determine what percentage of pilots predominately use the autopilot for flight. The results are displayed in Table 20 and Figure 21.

Table 20. Autopilot Usage Above 1000 Feet

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		1	2.5	2.5	2.5
	Strongly Agree	8	20.0	20.0	22.5
	Somewhat Agree	5	12.5	12.5	35.0
	Neutral	2	5.0	5.0	40.0
	Somewhat Disagree	13	32.5	32.5	72.5
	Strongly Disagree	11	27.5	27.5	100.0
	Total	40	100.0	100.0	

This statement had the most varied responses with a slight majority (56%) of the pilots either disagreeing or strongly disagreeing with the statement. Pilots agreed with the statement 32.5% of the time.

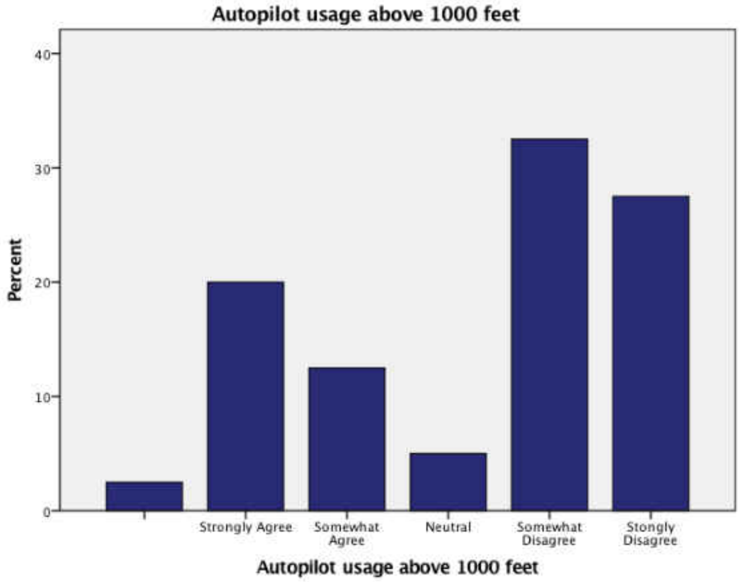


Figure 21. Autopilot Usage

Flight Evaluation - Quantitative Analysis

A quantitative analysis was completed using SPSS on the high and low altitude profiles and the sub-factors that comprised each profile. The profiles were analyzed both individually and then collapsed together with the independent variable being training as described in the Methods Section. Each of the three groups (high, low, and combined) were analyzed using a one-way ANOVA. In addition to descriptive statistics, regression analysis was conducted on the factors that made up each individual scenario score. Finally, the combined group was compared to the FAA standards for Airline Transport Pilot (ATP) certification. The results are discussed below.

High Altitude

Crews flew the high altitude scenario 20 times. There were nine untrained crews and 11 crews received the training as described in the Methods Section. SPSS was used to model a one-way ANOVA testing for the effects of training (independent variable) on the overall scenario event score. The descriptive statistics are summarized in Table 21.

Table 21. Descriptive Statistics High Altitude

High Altitude Scenario

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min-imum	Max-imum	Between-Compon-ent Variance
						Lower Bound	Upper Bound			
No		9	2.666	.70711	.23570	2.1231	3.2102	2.00	4.00	
Yes		11	3.727	.64667	.19498	3.2928	4.1617	2.00	4.00	
Total		20	3.250	.85070	.19022	2.8519	3.6481	2.00	4.00	
Mod-el	Fixed Effects			.67420	.15076	2.9333	3.5667			
	Random Effects				.53252	-3.5163	10.0163			.51653

The untrained crews had a mean score of 2.67 and a standard deviation of .70, which is slightly below the standard for FAA certification (a score of 3), while the trained crews had a mean score of 3.72 and a standard deviation of .65 which is above the FAA standard. There were no crews that received a score of one (1) indicating loss of control of the aircraft. There were also no scores of five (5) indicating a near perfectly flown scenario set.

A Levene's test was conducted on the high altitude scenario. The test indicates that there was no significant difference in the trained and untrained group variances. The test results were *F*(1, 18) = .674, *p* = .422. See Appendix E for additional detailed statistical test results.

A one-way ANOVA was conducted comparing the trained and untrained groups (see Table 22). There was a significant effect of training on the high altitude scenario score, *F*(1, 18) = 12.25, *p* = .003. The test confirms the research hypothesis that targeted training can be successful in helping pilots maintain aircraft control during an unusual and sudden startle event. In the case of the high altitude scenario, the null hypothesis is rejected.

Table 22. Tests of Between-Subjects Effects High Altitude

Dependent Variable: High Altitude Scenario

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	5.568[a]	1	5.568	12.250	.003	.405	12.250	.911
Intercept	202.368	1	202.368	445.210	.000	.961	445.210	1.000
CrewTrng	5.568	1	5.568	12.250	.003	.405	12.250	.911
Error	8.182	18	.455					
Total	225.000	20						
Corrected Total	13.750	19						

a. R Squared = .405 (Adjusted R Squared = .372) b. Computed using alpha = .05

The effect size was calculated from the ANOVA results using the formula $R^2=SS_M/SS_T$ with the result being n = .64. This represents a large effect size $R^2 = \eta^2$ =5.568/13.750 Eta η= .64. The result indicates that the average person in the experimental group would score higher than 73% of a control group that was initially equivalent (Coe, R., 2002).

Regression High Altitude

Regression analysis was conducted on the factors that made up the overall scenario event score. This was done to explore any significant factors leading to the overall score and to pinpoint possible areas of future training. The factors are summarized in the table below. Each factor was evaluated 20 times.

Table 23. Descriptive Statistics Regression High Altitude

	Mean	Std. Deviation	N
High Altitude Scenario	3.2500	.85070	20
Crew Training Received	.5500	.51042	20
Problem diagnosis	3.6000	1.14248	20
Pitch control	3.5000	1.10024	20
Roll control	3.8500	.81273	20
Altitude control	3.3500	.93330	20

The regression analysis was conducted in two blocks with block one being crew training and block two consisting of the described factors in the previous table. The ANOVA indicates that both models significantly improve the ability to predict the outcome variable compared to not fitting the model. Model 1 had *F*(1, 18) = 12.25, p = .003 and Model 2 had *F*(5, 14) = 10.02, p = .00. See Appendix E.

The regression coefficients were also analyzed to determine which factors (other than crew training in model 1) showed significance. Significance was noted for the factor of problem diagnosis p = .00. The other factors did not show significance (p > .05). Collinearity analysis indicates that there are no examples of multicollinearity (VIF > 10). See Appendix E

Low Altitude Scenario

The low altitude, low fuel scenario was flown by 20 crews. There were 10 trained and 10 untrained crews as described in the methods section. SPSS was used to analyze the results using a one-way ANOVA. The mean score of the untrained crews was 2.60 with a standard deviation of .70 and the trained crews was 3.70 with a standard deviation of .82. As with the high altitude scenario, the untrained crews performed below the ATP standards and the trained group performed above the standard. There were no scores of one (1) which would indicate a loss of control or crash of the aircraft. There were scores of five (5) indicating performance well above the FAA certification standard. The descriptive statistics are summarized in the Table 24 below.

Table 24. Descriptives Low Altitude

Low Altitude Scenario

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum	Between-Component Variance
						Lower Bound	Upper Bound			
No		10	2.600	.69921	.22111	2.0998	3.1002	2.00	4.00	
Yes		10	3.700	.82327	.26034	3.1111	4.2889	3.00	5.00	
Total		20	3.150	.93330	.20869	2.7132	3.5868	2.00	5.00	
Model	Fixed Effects			.76376	.17078	2.7912	3.5088			
	Random Effects				.55000	-3.8384	10.1384			.54667

A Levene's test was conducted on the low altitude scenario. The test indicates that there was no significant difference in the trained and untrained group variances. The test results were *F*(1, 18) = .450, *p* = ..511. See Appendix E.

A one-way ANOVA was conducted on the low altitude scenario to test the main effect of crew training. There was a significant effect of training on the low altitude scenario score, *F*(1, 18) = 10.37, *p* = .005. The test confirms the research hypothesis that targeted training be successful in helping pilots maintain aircraft control during an unusual and sudden startle event. In the case of the low altitude scenario, the null hypothesis is rejected. The ANOVA results are summarized in Table 25.

Table 25. Tests of Between-Subjects Effects Low Altitude

Dependent Variable: Low Altitude Scenario

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	6.050[a]	1	6.050	10.371	.005	.366	10.371	.861
Intercept	198.450	1	198.450	340.200	.000	.950	340.200	1.000
CrewTrng	6.050	1	6.050	10.371	.005	.366	10.371	.861
Error	10.500	18	.583					
Total	215.000	20						
Corrected Total	16.550	19						

a. R Squared = .366 (Adjusted R Squared = .330)

b. Computed using alpha = .05

The effect size was calculated from the ANOVA results using the formula $R^2=SS_M/SS_T$ with the result being $\eta = .60$. This represents a large effect size $R^2 = \eta^2 = 6.05/16.55$ Eta $\eta = .60$. The result indicates that the average person in the experimental group would, as in the high altitude scenario, score higher than 73% of a control group that was initially equivalent (Coe, 2002).

Regression Low Altitude

Regression analysis was conducted on the factors that made up the overall scenario event score, similar to the high altitude scenario. This was done to explore any significant factors leading to the overall score and to pinpoint possible areas of future training. The factors are summarized in the table below. Each factor was evaluated 20 times. See Table 26.

Table 26. Descriptive Statistics Low Altitude

	Mean	Std. Deviation	N
Low Altitude Scenario	3.1500	.93330	20
Crew Training Received	.5000	.51299	20
Missed approach	2.8000	1.10501	20
Checklist procedures	3.1500	.98809	20
Time Management	3.1500	1.03999	20
Fuel Management	3.2500	.91047	20
Approach and landing	3.3000	.92338	20

The regression analysis was conducted in two blocks with block one being crew training and block two consisting of the above described factors. The ANOVA indicates that both models significantly improve the ability to predict the outcome variable compared to not fitting the model. Model 1 had *F*(1, 18) = 10.37, p = .005 and Model 2 had *F*(6, 13) = 28.13, p = .00. See Appendix E.

The regression coefficients were also analyzed to determine which factors (other than crew training in Model 1) showed significance. Significance was noted for the factor of missed approach p = .01 and time management p = .00. The other factors did not show significance (p > .05). Collinearity analysis indicates that there are no examples of multicollinearity (VIF > 10). See Appendix E.

Each predictor had variance loading onto a different dimension. This also indicates no issues with multicollinearity.

Low and High Altitude Combined

The final set of analyses were conducted by collapsing the effect of crew training across both the low altitude and high altitude scenarios. There was no regression on the sub-factors of the combined scores due to the factors already being analyzed in the individual scenarios. Analysis was also conducted on the effects of pilot flying (PF), pilot monitoring (PM) and the type of aircraft involved (narrow body or wide body). Further analysis was conducted to compare both the high altitude and low altitude scenario vs the FAA standard for ATP certification for both the trained and untrained groups.

A total of 40 crews (80 individuals) volunteered for the study, of which 19 did not receive startle training and 21 received startle training. The mean for the untrained crews was 2.58 with a standard deviation of .6. The mean was 3.71 with a standard deviation of .71 for the trained group. The determination of the scenario score was described in the Methods Section. Table 27 summarizes the group means and standard deviation.

Table 27. Descriptive Statistics Combined

Dependent Variable: Low and High Altitude Combined

Crew Training Received	Mean	Std. Deviation	N
No	2.5789	.60698	19
Yes	3.7143	.71714	21
Total	3.1750	.87376	40

A further breakdown was analyzed to determine if there were significant differences between the high and low scenarios when comparing trained and untrained crews. The results are summarized in Table 28.

Table 28. High and Low Altitude Mean Comparison

Crew Training Received		High Altitude Scenario	Low Altitude Scenario
No	Mean	2.6667	2.6000
	N	9.0	10.0
	Std. Deviation	.70711	.69921
Yes	Mean	3.7273	3.7000
	N	11.0	10.0
	Std. Deviation	.64667	.82327
Total	Mean	3.2500	3.1500
	N	20.0	20.0
	Std. Deviation	.85070	.93330

A t-test was utilized to test for significance between the mean of the trained crews for both the low and high altitude scenarios. The test did not show significance in either case (p = .89 and .91) See Table 29.

Table 29. Means Comparison with Training

	Test Value = 3.7/3.73					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
High to Low	.140	10	.892	.02727	-.4072	.4617
Low to HIgh	-.115	9	.911	-.03000	-.6189	.5589

The untrained groups were also analyzed for significance between the low and high altitude scenarios. The mean for the untrained also did not show any statistical significance with p = .78 and .76. See Table 30.

Table 30. Means Comparison No Training

	Test Value = 2.60/2.67					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
High to Low	.283	8	.784	.06667	-.4769	.6102
Low to High	-.317	9	.759	-.07000	-.5702	.4302

The results of the t-tests indicate that there was no significant difference between the trained and untrained groups when the trained and untrained groups are combined across scenario sets. The research question that sought to determine if the proximity of the scenario had any effect on the outcome was answered. The data indicates that the null hypothesis in this case is retained.

As with both the high and low altitude scenario, a Levene test for equal variances was conducted. The test did not yield significant results *F*(1, 38) = .046, p = .83, therefore the assumption is that the variances are equal across the groups. See Appendix E.

The next step in the analysis was an ANOVA calculated using SPSS. The ANOVA tested for crew training when collapsed across both high and low altitude scenarios. The test revealed that the effect of crew training was significant *F(*1, 38) = 28.89, p = .00. This means that the trained crews showed a statistically significant increase in performance due to the effect of crew training. The results are summarized in Table 31.

Table 31. Tests of Between-Subjects Effects Combined

Dependent Variable: Low and High Altitude Combined

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	12.858[a]	1	12.858	28.881	.000	.432	28.881	.999
Intercept	395.058	1	395.058	887.387	.000	.959	887.387	1.000
CrewTrng	12.858	1	12.858	28.881	.000	.432	28.881	.999
Error	16.917	38	.445					
Total	433.000	40						
Corrected Total	29.775	39						

a. R Squared = .432 (Adjusted R Squared = .417)

b. Computed using alpha =.05

The effect size was calculated from the ANOVA results using the formula $R^2=SS_M/SS_T$ with the result being η = .66. This represents a large effect size $R^2 = \eta^2$ =12.86/29.78 Eta η= .66. The combined effect size was larger than both the high and low altitude scenario effect sizes. The result indicates that the average person in the experimental group would score higher than 73% of the control group that was initially equivalent (Coe, 2002).

The next test measured for significant differences in the pilot flying, pilot monitoring (Captain or First Officer), and the type of aircraft. ANOVA testing for differences between PF, PM and NB/WB yielded no significant results. The results are summarized in Table 32.

Table 32. ANOVA Other Factors

		Sum of Squares	df	Mean Square	F	Sig.
Low and High Altitude Combined	Between Groups	12.858	1	12.858	28.881	.000
	Within Groups	16.917	38	.445		
	Total	29.775	39			
Captain or First Officer	Between Groups	.226	1	.226	.877	.355
	Within Groups	9.774	38	.257		
	Total	10.000	39			
Pilot Monitoring	Between Groups	1.684	1	1.684	2.410	.129
	Within Groups	25.158	36	.699		
	Total	26.842	37			
Narrow or Wide body acft	Between Groups	.095	1	.095	.367	.548
	Within Groups	9.880	38	.260		
	Total	9.975	39			

The ANOVA indicated that were no differences in scenario scores with either the Captain or First Officer flying *F*(1, 38) = .88, p=.36. It also indicated that there was no significance to which crew member was the pilot monitoring *F*(1, 36) = 2.4, p = .129. Finally, the test indicated no significance between aircraft type *F*(1, 38) = .37 p = .55. This test answered the research question seeking to explore if significant differences could be determined within these categories. In the above cases, the hypothesis of significant differences is rejected and the null hypothesis is retained.

Performance versus FAA Standard

The next segment of the analysis was conducted using a series of t-tests to explore differences of crew performance from the FAA standard for ATP certification. The crews for this study had just completed their recurrent training cycle where they were certified to FAA standards.

The first t-test explored the FAA standard of three (3) to the overall mean (high and low altitude scenario) of the untrained crews. The results are summarized in Table 33.

Table 33. T-test Training versus FAA Standard

	Test Value = 3					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
No Crew Training	-3.024	18	.007	-.42105	-.7136	-.1285
Crew Training	4.564	20	.000	.71429	.3878	1.0407

The t-test revealed, that in the case of no crew training, that performance was significantly below the FAA certification standards. The mean score difference was -.42 that resulted in a significance of p = .01. This indicates that crew performance during a startle event is significantly different from the FAA standard. In the case of crew training, the results were also significant but in a positive direction resulting in a mean score difference of .71 with a p = .000.

Survey and Performance Correlations

The final set of analysis was conducted to determine if survey responses correlated with crew performance. Analysis was conducted separately for both the trained and untrained crews. The first set of tests for the untrained crews used a Pearson's correlation to highlight significant result. The results are summarized in Appendix E.

There were no significant survey responses that correlated with the performance of the untrained crews. There were, however, significant correlations between two of the survey statements. The statements "I often hand-fly below 10,000 feet" and "I often practice my raw data skills" was significant with p = .02. The statement "I am comfortable flying with raw data" and "I often practice my raw data skills"' also showed a significant correlation with p = .03. These questions should be correlated because one question generally produces the other.

Significant correlations were also explored for the group that received crew training. The results were identical to the untrained group with the same two statements showing significant correlations p = .00 and p = .01. The results are also summarized in Appendix E.

Results Summary

The data was collected and analyzed in two distinct parts: the pilot survey, and the crews flying a scenario in the simulator. The survey was designed to gain the pilot's perspective on their experiences and attitudes towards flying. The profile scenarios that were flown in the simulator sought to produce a startle event for the crews and then record their performance.

The survey data indicated that over 80% of the pilots reported that they had incurred an event while flying that surprised them. In addition, a good mix of both civilian (67%) and military (33%) pilots participated in the study. The survey questions were not correlated with crew performance in either the low altitude or high altitude scenario and thus were not a predictor of crew performance.

The event scenarios were all flown in an FAA certified Level-D full flight simulator (FFS), with accurate visual and vestibular sensory input. In both scenario types, the crews that received training that consisted of a briefing and simulator practice showed a significant improvement in performance than the crews that did not receive training. This confirmed the hypothesis that targeted training on mitigation of startle effect could increase crew performance. Regression analysis was also conducted on the factors that made up each crew performance. The analysis suggests that problem recognition in the high altitude scenario and the missed approach in the low altitude scenario were significant predictors of performance on the overall profile.

Crew performance, collapsed across both scenarios, was also measured against the FAA standards for ATP certification. When crews received training, the data showed a significant improvement in performance vs the FAA standard. Data also indicated that untrained crews performed significantly worse than the FAA standard when presented with a startle event. This result was unexpected. There were no significant results when examining the pilot flying, pilot monitoring, or aircraft type as it related to crew performance, rejecting the hypothesis that these factors would be significant. There were also no significant results between the low altitude and high altitude scenarios when looking at both trained and untrained crews. The null hypothesis was retained for this research question related to event proximity (low or high altitude).

Finally, the survey responses were correlated with crew performance. There were no significant correlations between performance and survey responses. There were, however, two significant correlations that were solely related to the survey questions.

CHAPTER V

DISCUSSION

Introduction

Startle effect has been well documented for the past 50 years. Vlasak (1969) found significant impairment in cognitive function for the first 15 seconds following a startle event. Other studies have shown similar results. Unfortunately, it is often during this period that critical aircraft handling decisions must be made. According to Boeing (2012), inflight loss of control is the leading cause of airline fatalities. Recent accidents such as Air Asia and FlyDubai indicate the inflight loss of control continues to be a significant safety issue for airlines. This study sought to determine if targeted training could improve simulator performance of crews during a startle event. Volunteer airline pilots flew two different startle scenarios in a full flight simulator. The scenarios were designed to be similar to events that have caused major airline accidents. The volunteer groups were divided into low altitude scenario groups and high altitude scenario groups. The groups were further divided into trained and untrained groups. Data analysis was conducted on the main effect of training within and between each volunteer group. Additional analysis was also conducted on the sub-factors that made up each scenario such as pilot flying, aircraft type, and maneuver sub-parts.

Findings

The data showed that targeted startle training could improve crew performance while flying startle scenarios in the simulator. Significance was found for the trained crews in both the low and high altitude scenarios when compared to the untrained crews. The effect of the training was shown to be high, predicting that trained crews would perform 73% better than untrained crews (Coe, 2002). Trained crews also showed a significant increase in performance when compared to the FAA standards for ATP certification. The results answered the research question asking whether targeted training could increase crew performance during a startle event in the simulator.

In the high altitude scenario, crews were exposed to failures similar to what occurred in the Air France 447 accident as described in the literature review (BAE, 2012). Crews that received training that consisted of a briefing and simulator practice, on how to handle startle events. This group performed significantly better than the FAA standard and significantly better than the crews that flew the same profile but did not receive training.

The low altitude scenario was modeled after the Colgan accident in Buffalo, NY, and presented the crews with a low altitude startle event that, in most cases, pushed the crews into a missed approach in a low fuel situation. The results were similar to those of the high altitude profile in that trained crews performed significantly better than the untrained crews.

Unexpected results were found in the crews that did not receive the startle training when compared to the FAA standard. The data analysis showed that the crews performed statistically significantly below the level for ATP certification for at least a portion of the time during the maneuver profile. It should be noted that there were no crews who lost control of the aircraft during the profile and that all crews eventually had a successful outcome. This data explored the initial reaction of the crews since this is the critical decision making time frame. The crews for this study had just completed their annual recurrent training cycle for their respective aircraft under what is known as the advanced qualification process (AQP). In an AQP, Training Program, when a crew (or individual pilot) falls below the FAA standard they receive training and then are required to perform the skill again. This concept is termed “train to proficiency.” Crews in this research study were only presented the scenario one time. This “first look” only takes a snapshot of a crews’ performance in time and indicates where training could be effective. It is not meant to be extrapolated to overall crew competency

Each maneuver scenario was made up of several sub-factors or components that comprised the overall score. These factors were analyzed to determine their significance in making up the total score and to uncover possible dimensions where training should be targeted. The results for the high altitude scenario indicated that the most significant factor in determining scenario success was “problem identification”. This was consistent with previous research findings which showed that when crews make an initial wrong decision, the in-flight issue tends to rapidly degrade. The low altitude scenario was somewhat less clear in significant factors. Time management was a significant predictor of crew performance. The missed approach was also a significant predictor which was unexpected. If crews performed the missed approach sucessfully, the rest of the scenario generally was graded better than if the missed approach was incorrectly flown. Minimum amount of fuel (45 minutes) at start of the missed approach likely influenced this result.

The research also sought to determine if the pilot flying (Captain or First Officer) resulted in significant differences in maneuver performance. The study's data demonstrated that crew performance was not affected by which pilot was flying. The study looked at this dimension since in a majority of airline accidents/incidents, the Captain is the pilot flying and it is the first flight of the trip pairing (United, 2016).

Simulator data also did not uncover any significance between the types of aircraft in predicting the success of the scenario. Aircraft were grouped into narrow body and wide body categories. Narrow body aircraft pilots have a greater frequency of takeoffs and landings than those pilot who fly wide-body aircraft. This difference in frequency may add to narrow body aircraft pilots flying proficiency and increase the success of responding to a startle event. In addition, most wide-body aircraft rely heavily on automation due to the long duration of their flights possible making a sudden startle event more challenging. However, pilots who fly wide-body aircraft generally have a more experienced background than pilots who fly narrow-body aircraft due to the airlines' seniority system. The research question on practice and experience sought to determine if practice and/or experience could influence the overall maneuver score. The research question relating to aircraft type did not show it affected the maneuver scenario in either a positive or negative way. There was no significant difference in either pilot group.

The pilots in this research study were also requested to complete a survey that asked questions about their background and flying preferences. The survey responses were not linked to any crew member's individual performance, but were analyzed in aggregate with regard to simulator performance. The pilot flying the maneuver scenario was the only crew member asked to complete the survey. This was done to keep the survey responses equal to the number of crews observed (40) and to compare the responses with the pilot who flew the simulator profile. The survey responses indicated that the pilots generally hand flew the airplane below 10,000 feet and that they knew the proper pitch and power for various phases of flight. Furthermore, a significant portion of the pilots surveyed indicated that they had received formal aerobatic training. The responses given do not necessarily correlate with simulator performance when taken as a whole. If the trained crews were the only ones examined, then the survey responses correlate with positive performance; however, when untrained crews were added, pilots tended to overestimate their flying performance.

Significance

The data recorded for this study showed that targeted training can help pilots bridge the cognitive gap when startled. Crews performed equally well in both the high altitude and low altitude scenario, suggesting that the training had a broad array of effectiveness. Both scenarios recorded a similar main effect: power of eta = .6

suggested a medium to large effect size. The training offered consisted of a briefing to explain the effects of being startled along with a short and simple procedure to help mitigate the startle effect and regain (or keep) control of the aircraft. The study was not designed to eliminate the startle response which would be very difficult to accomplish, but sought to help crews manage the period of cognitive impairment. In summarizing the training, the motto “live for the next 60 seconds” was often used. This is the time in which the most cognitive impairment occurs. Unfortunately, crews often have to make critical decisions in this time to keep control of the aircraft. The data indicated that trained crews were more successful in managing the aircraft following a startle event than those crews that did not receive training. Crews that received training flew significantly better than the FAA standards for ATP certification, indicating a positive shift in event handling even versus a standard crew.

The training suggested in this study has implications for the airline industry as a whole. As previously stated, crews that were not trained showed a statistically significant degradation below FAA ATP standards. Following the startle event, the untrained crews lapsed out of ATP standards as described in the Methods Section of this study. All of the crews were eventually able to successfully recover from the simulated situation, however it is the decision making at the onset which can prove critical to event outcome. During the study there were no crews put the aircraft into an undesired aircraft state (UAS). This suggests that current airline training may be improved by incorporating startle training. Several published papers allude to this idea in that airline training has become rote and routine; not challenging crews with new situations and scenarios that expand flying knowledge and experience (Casner, Geven, & Willliams, 2012). Most airlines have a standard training profile that is determined by the regulatory requirements of the FAA. This training is generally the same from year to year resulting in repetition and expected outcomes. Training outside of this set standard is often referred to as "proficiency training" (United Airlines, 2016), and usually exposes the crew only at predetermined cycles and has more to do with technical failures and not cognitive loading.

The pilots, through their survey responses, indicated that they generally hand fly the airplane below 10,000 feet. Hand flying below this altitude provides a boost to skill maintenance due to the fact that changes in all phases of flight occur frequently. Takeoffs, approaches, and landings all require changes to aircraft speed, configuration, and navigation (lateral and vertical). These maneuvers challenge piloting skills and keep them sharp (Gillen, 2014). Overall piloting skills may be a key element in aircraft control. A pilot proficient in hand flying will require less cognitive resources (to fly the airplane) and may be able to devote more cognitive processes to problem detection. Pilots in the United States generally hand fly the aircraft more than in other parts of the world. In a paper presented at the Lufthansa Human Factors Conference (Gillen, 2014), pilots employed by airlines from various global carriers were surveyed about their hand flying practices. The results are summarized in Table 34.

Table 34. Hand Flying Preferences

Company Description	Company Policy	Actual Practice
United States - Global	Company policy states that the automation level is at the discretion of the Captain.	A majority of pilots hand fly the aircraft extensively below 10,000 feet.
Major European	Hand flying is encouraged to maintain proficiency	Most pilots report hand flying below 10,000 feet.
Middle East	Company policy prohibits hand flying above 10,000 feet.	Pilots report that they generally engage the autopilot at 1,000 feet on departure and disengage on approach once the aircraft is fully configured for landing.
Asia	Company policy encourages hand flying to increase pilot proficiency	Company regularly uses FOQA data in disciplinary cases against pilots. As a result pilots rarely hand fly the aircraft.
Southeast Asia	Company policy states that automation is at the discretion of the Captain	Manual flying varied widely depending on the flight crews.

Lack of flying skills become more apparent when system failures cause pilots to revert to manual flying skills to maneuver the aircraft. Simple failures can lead to a cascade of errors and pilot confusion that in turn can lead to an undesired aircraft state (Gillen, 2014). These system failures can also tax a pilot's cognitive resources well beyond their ability to cope with the situation (Gillen, 2014). Based on hand flying preferences, the pilots in this study should be considered the most proficient and thus the results should trend higher when compared to pilots who hand fly less.

Critical Evaluation

Studies involving airline crews in the United States are often difficult to complete. Airlines and their respective pilot unions are reluctant to have a researcher record live data on a crews' performance. The only way to obtain permission for such a study is to have the data de-identified so that no individual performance can be linked back to a specific pilot or crew. As such, this research was only able to observe volunteer crews one time to determine the effect of the training. Training results could have been more conclusive if a revisit of the trained crews had taken place. Unfortunately, this could not occur as it would make the individual pilots identified which would violate the permission letter from the participating airline(s). Data suggests that unused training skills will decrease in effectiveness over time and that deliberate practice is required. It is the opinion of the researcher that if the training presented in this study is not practiced, then the effectiveness (of the training) will most likely decrease over time.

Any type of simulator training could be reasonably expected to an increase in the crews' overall scenario scores. Study subjects were asked to volunteer immediately following their annual recurrent training cycle which included 8-12 hours of simulator training/checking. All crews involved in the study received simulator training and practice immediately preceding participation in this study. As a result, scores may be skewed higher than if crews flew the maneuvers without any previous practice.

Results in this study could possibly be biased in a positive direction due to the voluntary nature of the participants. Generally, pilots who volunteer for these types of studies are comfortable in their flying skills and interested in aviation safety. Pilots who have difficulty in training generally would choose to not participate. Therefore, the overall results might be skewed towards the higher end than the average airline pilot population.

Implications

The startle effect is not a new concept and the effects of being startled are well known. What is not well known is how to mitigate the startle effect in airline crews where critical decision making must often take place concurrent with the time of cognitive impairment following a startle event. Training and practice have been shown to increase a pilot’s response to aircraft control during an event that catches a crew unexpectedly. Targeted training should be procedural in nature and seek to become skill base (best) or rule based behavior. This method requires a consistent and systematic approach to dealing with unusual events.

To be effective, training that is described in this study should be implemented in both initial and recurrent pilot training in addition to being reinforced in actual line flying. This study has shown that significant positive effects of training can be realized in as little time as 60 minutes.

Follow on studies should look at the effectiveness of this type of training over longer intervals. The training presented in this study was designed to be broad in nature and cover various states of contingencies as it relates to a startle event. Such training is intended to be applicable in a general way and is not intended to be aircraft or airline specific. It is more of a philosophy in dealing with unusual events at the initial decision making point to help bias a successful outcome. In the Air France and Colgan accidents, that were described in the literature review, aircraft control was lost in the first 30 seconds following a startle event. Training should focus on this time period to be most effective. Although sudden and unusual events cannot be prevented in aviation, a pilot's response to them (especially at the onset) can be positively influenced to aid in a successful outcome. There is not a single solution in airline training to eliminate the risk of a startle event, only mitigating factors, that when presented in multiple layers serve to aid crews in successfully handling the event.

Recommendations

Startle training should be added to the training programs at airlines to make crews aware of the effects on performance of being startled and mitigation strategies that can help pilots successfully fly the aircraft immediately following a startle event. Startle training, in order to be effective, has to be reinforced at specific training intervals such as during each pilot's initial and annual recurrent training cycle. Positive results were shown in this study where training consisted of both classroom and simulator practice lasting approximately one hour. Training should focus on what happens from a cognitive standpoint and what steps pilots should take to stabilize the aircraft so that they can then determine the course of action to safely fly the aircraft. To be effective, this training will have to be varied to prevent habituation.

Recommendations for Further Research

Research into startle training should continue. This research should attempt to identify the best training interval for startle training to prevent degradation of startle response skills. There may be a link between hand flying ability, and the general ability to handle a startle response. In cases such as the Air France and Colgan accidents, the pilots were faced with a sudden event that forced them into hand flying the aircraft. Pilots who are competent in hand flying require less cognitive resources to do so, and may be able to devote more resources to problem definition. This would be a good area for future research as well. One further area for research may look at the link between often practiced unusual situations in the simulator and their negative transfer to unrelated events in the actual aircraft. During training, most of the maneuvers are performed at low altitudes and require an immediate pitch up of the aircraft's nose (engine failures, windshear, and missed approaches are all examples). While these responses are appropriate for low altitudes, the initial response to pitch the nose up may not be appropriate at high altitudes such as seen in the Air France accident.

Conclusion

This study showed that targeted training can improve crew performance in the simulator during a startle event. Given that skills learned in the simulator are generally well transferred to actual operations, the study results should also be highly transferrable to the actual aircraft. Data shows that increased startle training could significantly improve a pilot’s reaction to a startle event.

Startle events continue to be a major trigger resulting in aircraft inflight loss of control. Although not every event will result in a loss of aircraft control, training can help bridge the cognitive gap that exists during the initial seconds of a startle event. Training such as what was presented in this study should be added to airline training programs to aid crews towards a successful outcome of a startle event. A key element in dealing with a startle event often involves manual manipulation of the aircraft controls. Pilots who are proficient in hand flying will have an advantage in dealing with a startle event. Training, practice, and hand flying each hold an important element in successfully mitigating a startle event and preventing an inflight loss of control. Further studies should seek to determine the optimum integration of these three elements. It will take a partnership between the airlines, pilots, and the regulators to implement startle training. Such training can be a key mitigation strategy in reducing the leading cause or airline fatalities.

APPENDICES

Appendix A

Definitions

AC: Advisory Circular

ACO: Aircraft Certification Office

AD: Airworthiness Directive

AEG: Aircraft Evaluation Group

ALPA: Airline Pilots Association

AQP: Advanced Qualification Program

ARAC: Aviation Rulemaking Advisory Committee

ASAP: Aviation Safety/Accident Prevention

ASRS: Aviation Safety Reporting System

ATA: Air Transport Association of America

ATC: Air Traffic Control

ATIS: Automatic Terminal Information Service

ATP: Airline Transport Pilot

ATS: Air Traffic Services

BIS: Basic Instrument Skills: The ability to fly the aircraft solely by reference to the raw data without the use of auto-throttles, flight director, or map mode.

CFIT: Controlled Flight into Terrain

CMO: Certificate Management Office

CRM: Crew Resource Management

FAA: Federal Aviation Administration

FAR: Federal Aviation Regulations

FCOM: Flight Crew Operating Manual

FCU: Flight Control Unit

FMS: Flight Management System

FOEB: Flight Operations Evaluation Board

FSB: Flight Standardization Board

FSDO: Flight Standards District Office

GPS: Global Positioning System

GPWS: Ground Proximity Warning System

HF: Human Factors

ICAO: International Civil Aviation Organization

IFR: Instrument Flight Rules

IOE: Initial Operational Experience

ILS: Instrument Landing System

JAA: Joint Aviation Authorities

JAR: Joint Aviation Requirements

LNAV: Lateral Navigation

LOFT: Line Oriented Flight Training

LOS: Line Operational Simulations

Modern Aircraft/Glass Aircraft: Aircraft that have advanced automation to include: CAT III capability, auto-throttles, flight director, FMC, and CRT displays instead of actual instruments, the ability to LNAV and VNAV

NASA: National Aeronautics and Space Administration

NOAA: National Oceanic and Atmospheric Administration

NOTAM: Notice to Airmen

NTSB: National Transportation Safety Board

PDC: Pre-Departure Clearance

PF: Pilot Flying

PFD: Primary Flight Display

PM: Pilot Monitoring

PTS: Practical Test Standards Defined by the FAA Pilot Qualification.

RNP: Required Navigation Performance

TCAS: Traffic Alert and Collision Avoidance System

VNAV: Vertical Navigation

VOR: Very High Frequency Omnidirectional Radio Range

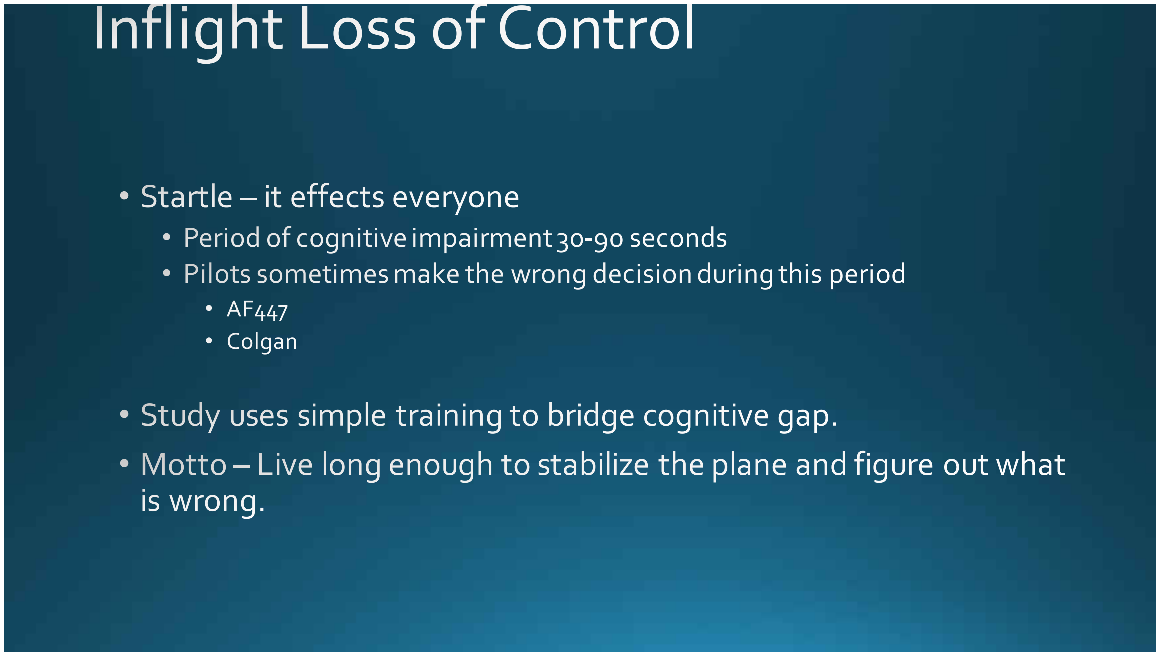
Appendix B

Briefing Materials

The following are slides from a Power Point presentation. The crews that received training were shown these slides as part of the classroom briefing.

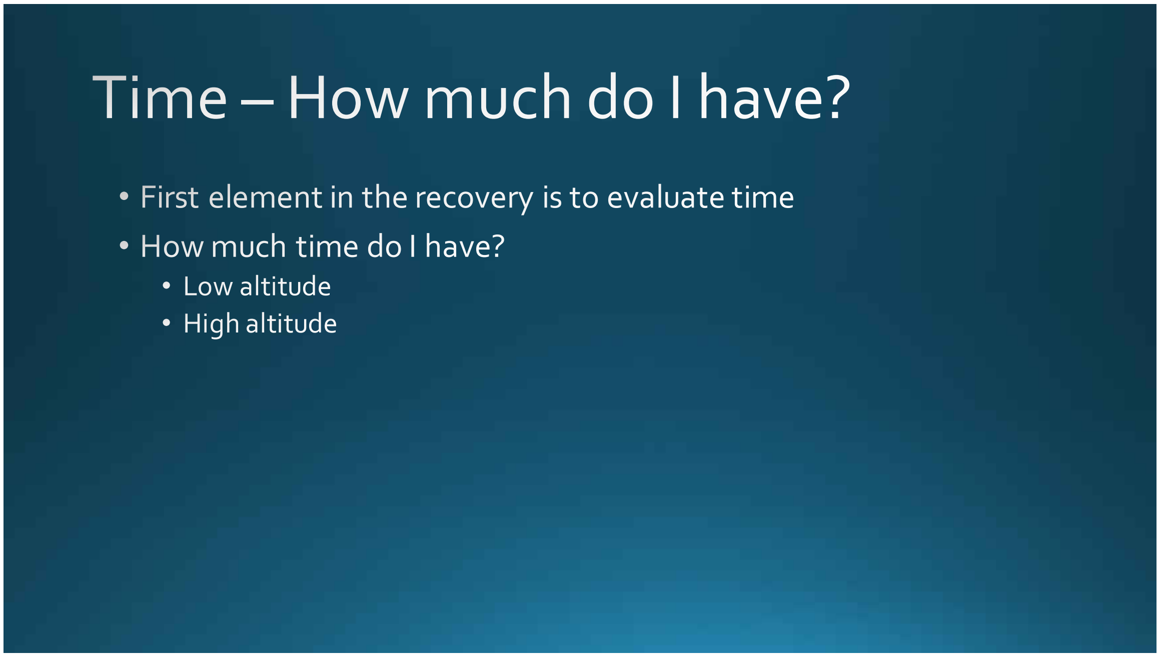
Startle Effect

Training



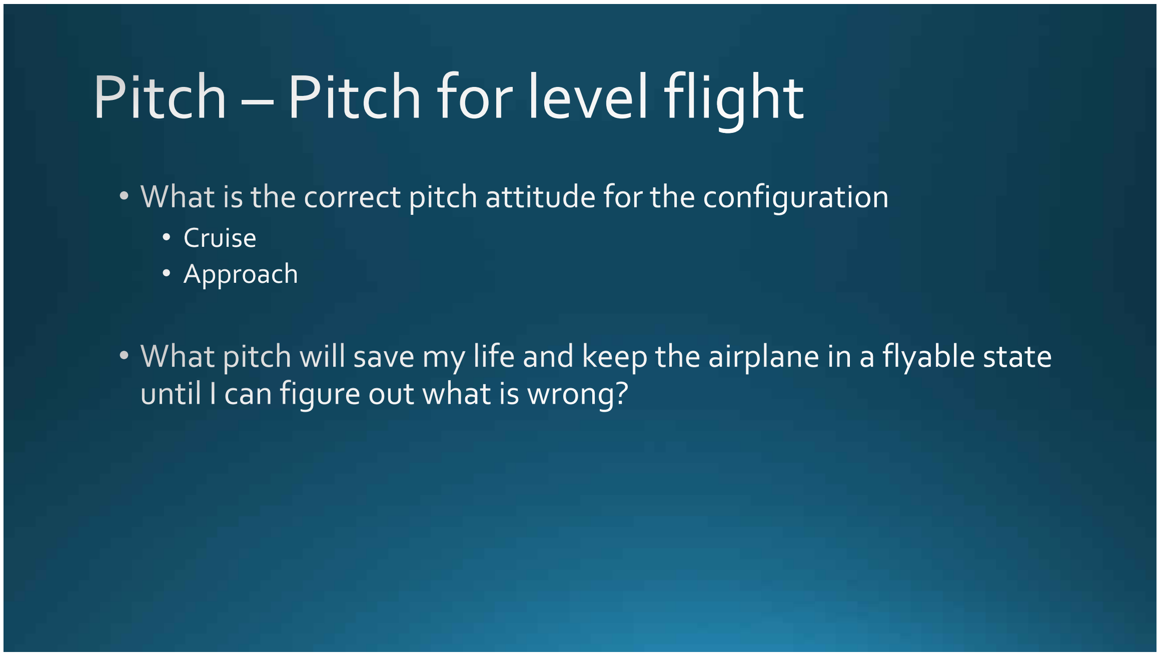
Inflight Loss of Control

- Startle – it effects everyone
  - Period of cognitive impairment 30-90 seconds
  - Pilots sometimes make the wrong decision during this period
    - AF447
    - Colgan
- Study uses simple training to bridge cognitive gap.
- Motto – Live long enough to stabilize the plane and figure out what is wrong.



Time – How much do I have?

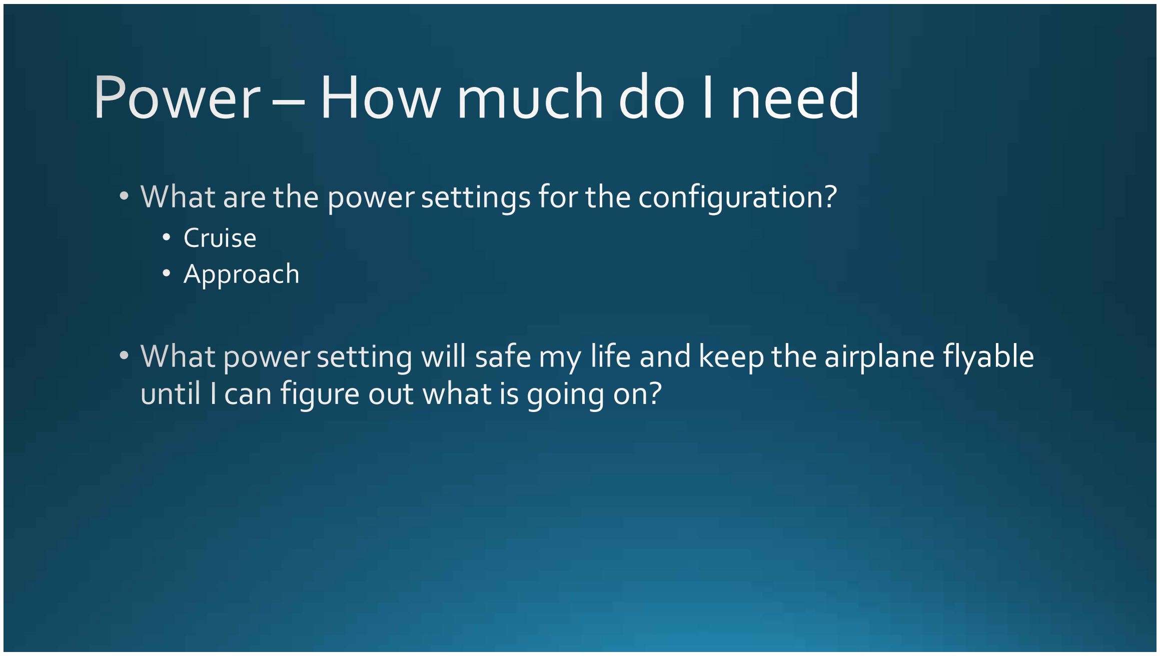
- First element in the recovery is to evaluate time
- How much time do I have?
  - Low altitude
  - High altitude



Pitch – Pitch for level flight

- What is the correct pitch attitude for the configuration
  - Cruise
  - Approach

- What pitch will save my life and keep the airplane in a flyable state until I can figure out what is wrong?



Power – How much do I need

- What are the power settings for the configuration?
  - Cruise
  - Approach

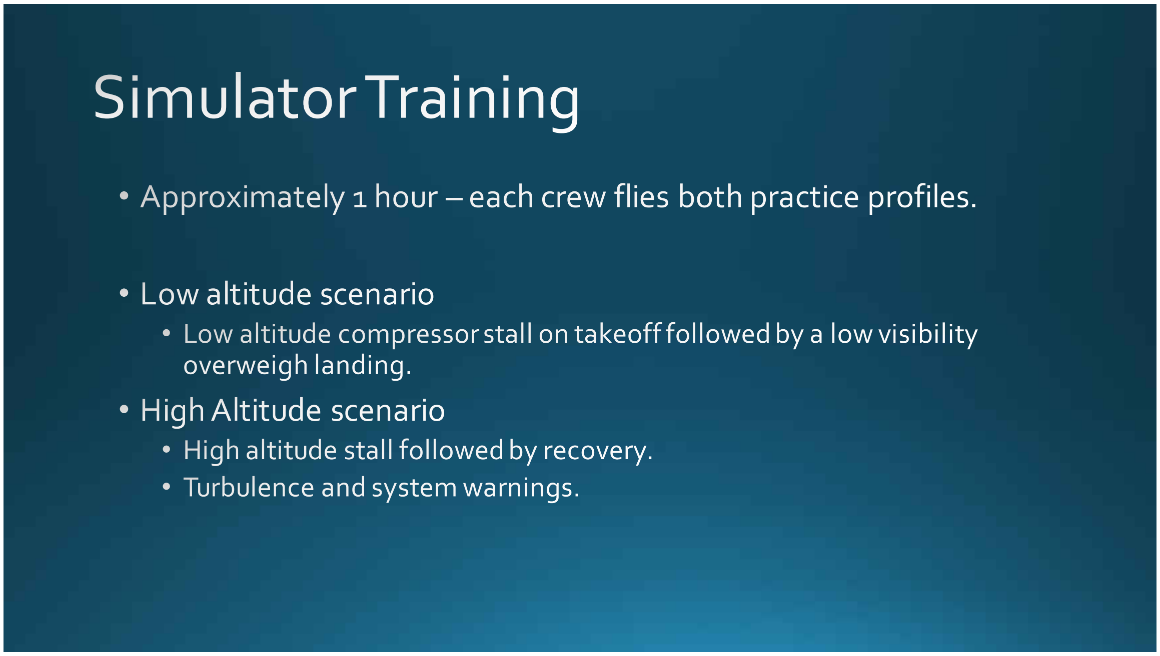
- What power setting will safe my life and keep the airplane flyable until I can figure out what is going on?

Bank – Level the wings

- Wings level if terrain not an issue.
- Fly straight until you can figure out your issue.

Training Mnemonic – Say it out loud

- T – Time – how much do I have
- A – Attitude (pitch and bank)
  - Pitch 3-5 deg
  - Bank – as level as you can
- P – Power
  - Set 80-85%



Simulator Training

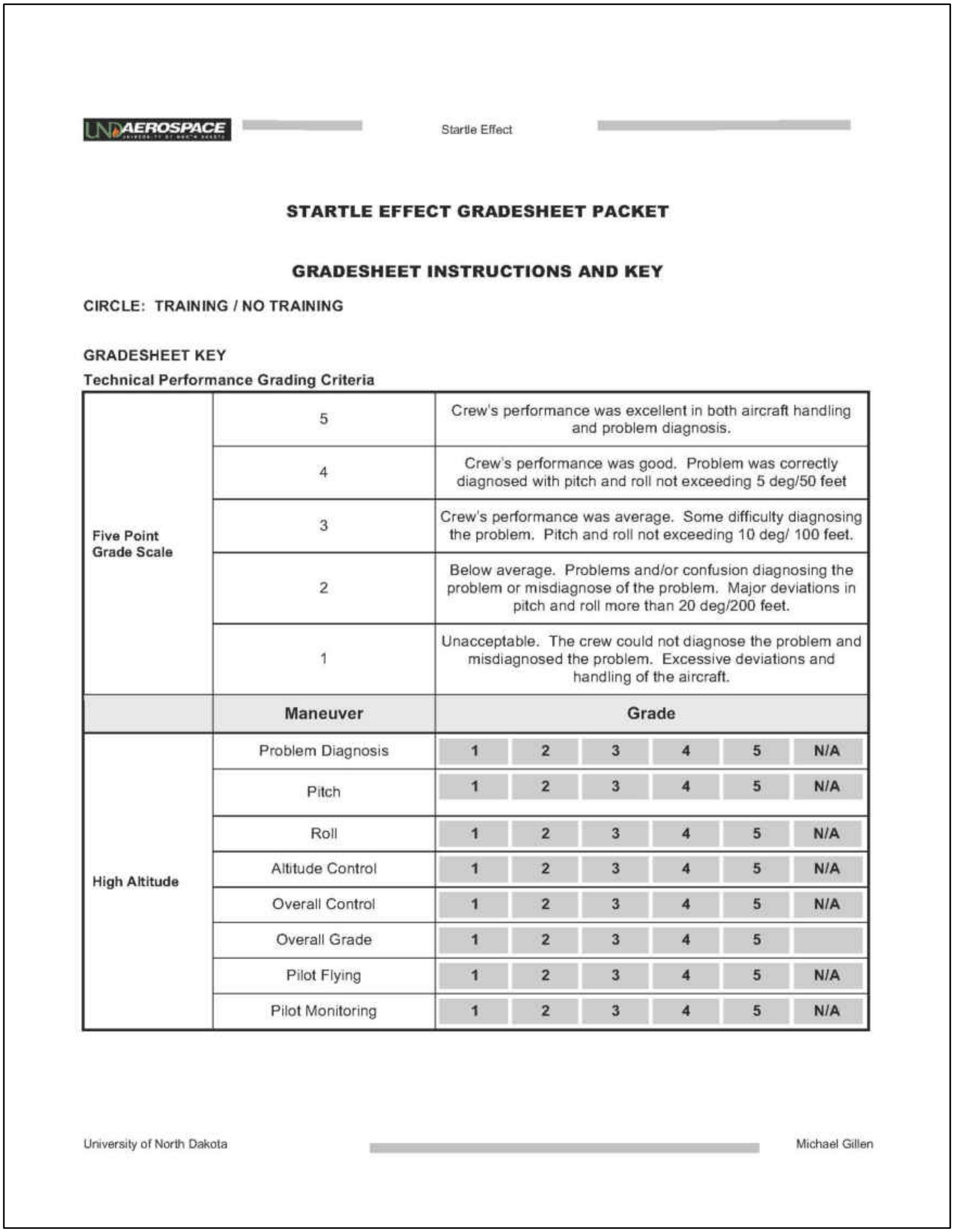
- Approximately 1 hour – each crew flies both practice profiles.

- Low altitude scenario
  - Low altitude compressor stall on takeoff followed by a low visibility overweigh landing.
- High Altitude scenario
  - High altitude stall followed by recovery.
  - Turbulence and system warnings.

Appendix C

Grade Sheets and Survey

Note: The grade sheets and survey are presented as they were used by the investigator during the data collection process. They have only been formatted to fit within the page margins of this paper.



Startle Effect

STARTLE EFFECT GRADESHEET PACKET

GRADESHEET INSTRUCTIONS AND KEY

CIRCLE: TRAINING / NO TRAINING

GRADESHEET KEY

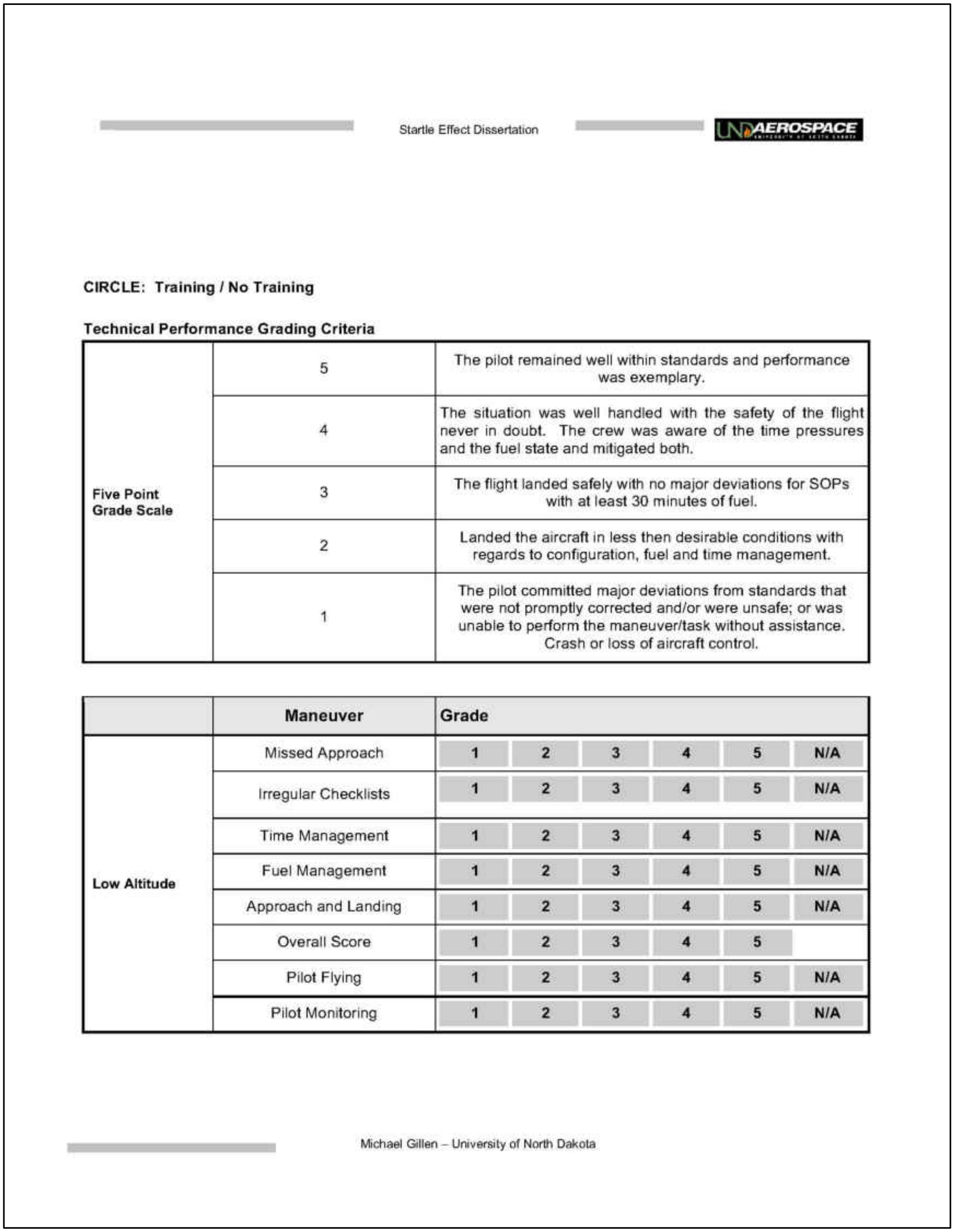
Technical Performance Grading Criteria

Five Point Grade Scale	5	Crew's performance was excellent in both aircraft handling and problem diagnosis.
	4	Crew's performance was good. Problem was correctly diagnosed with pitch and roll not exceeding 5 deg/50 feet
	3	Crew's performance was average. Some difficulty diagnosing the problem. Pitch and roll not exceeding 10 deg/ 100 feet.
	2	Below average. Problems and/or confusion diagnosing the problem or misdiagnose of the problem. Major deviations in pitch and roll more than 20 deg/200 feet.
	1	Unacceptable. The crew could not diagnose the problem and misdiagnosed the problem. Excessive deviations and handling of the aircraft.

	Maneuver	**Grade**					
High Altitude	Problem Diagnosis	1	2	3	4	5	N/A
	Pitch	1	2	3	4	5	N/A
	Roll	1	2	3	4	5	N/A
	Altitude Control	1	2	3	4	5	N/A
	Overall Control	1	2	3	4	5	N/A
	Overall Grade	1	2	3	4	5	
	Pilot Flying	1	2	3	4	5	N/A
	Pilot Monitoring	1	2	3	4	5	N/A

University of North Dakota

Michael Gillen

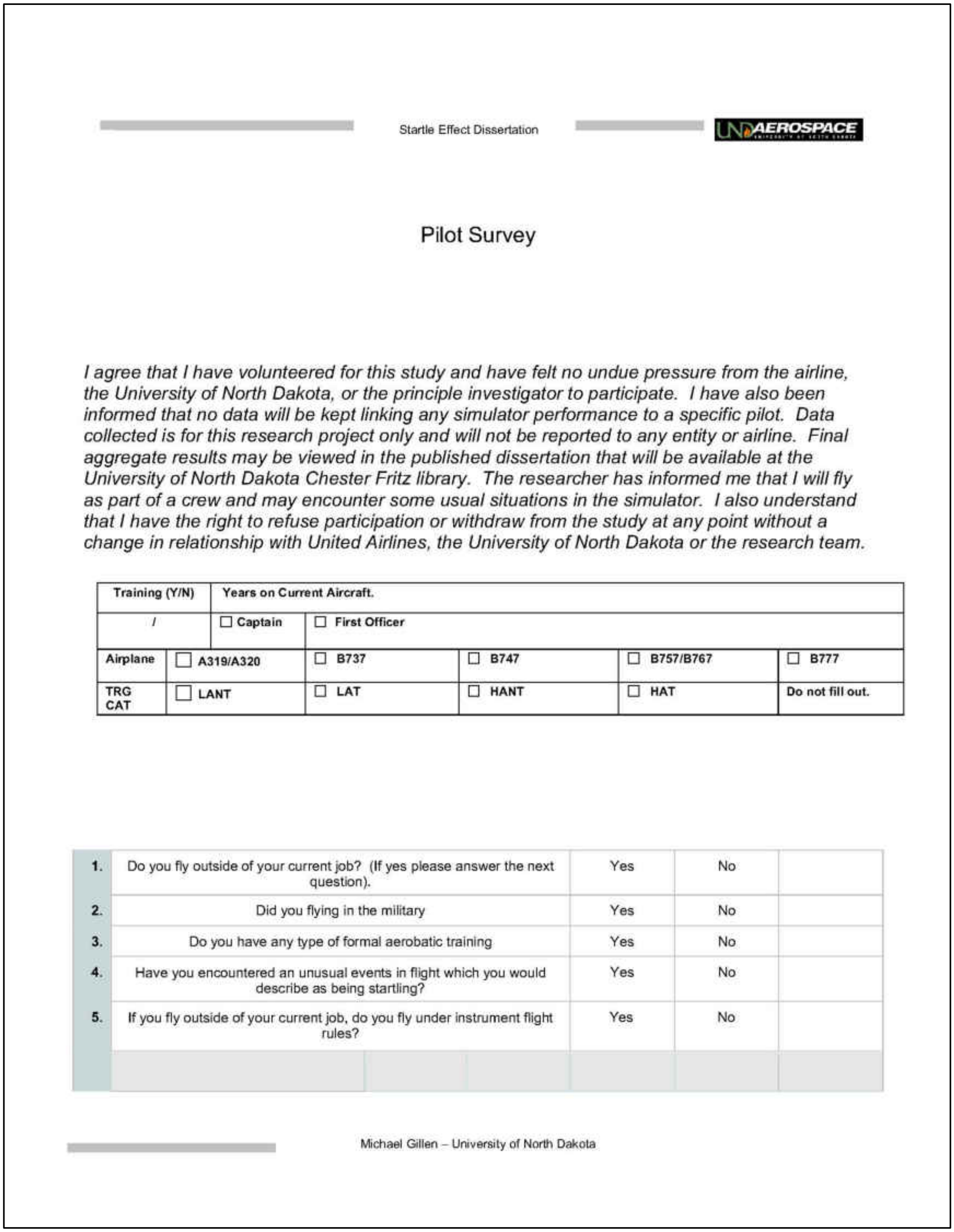


CIRCLE: Training / No Training

Technical Performance Grading Criteria

Five Point Grade Scale	5	The pilot remained well within standards and performance was exemplary.
	4	The situation was well handled with the safety of the flight never in doubt. The crew was aware of the time pressures and the fuel state and mitigated both.
	3	The flight landed safely with no major deviations for SOPs with at least 30 minutes of fuel.
	2	Landed the aircraft in less then desirable conditions with regards to configuration, fuel and time management.
	1	The pilot committed major deviations from standards that were not promptly corrected and/or were unsafe; or was unable to perform the maneuver/task without assistance. Crash or loss of aircraft control.

	Maneuver	**Grade**					
Low Altitude	Missed Approach	**1**	**2**	**3**	**4**	**5**	**N/A**
	Irregular Checklists	**1**	**2**	**3**	**4**	**5**	**N/A**
	Time Management	**1**	**2**	**3**	**4**	**5**	**N/A**
	Fuel Management	**1**	**2**	**3**	**4**	**5**	**N/A**
	Approach and Landing	**1**	**2**	**3**	**4**	**5**	**N/A**
	Overall Score	**1**	**2**	**3**	**4**	**5**	
	Pilot Flying	**1**	**2**	**3**	**4**	**5**	**N/A**
	Pilot Monitoring	**1**	**2**	**3**	**4**	**5**	**N/A**



Startle Effect Dissertation

UND AEROSPACE

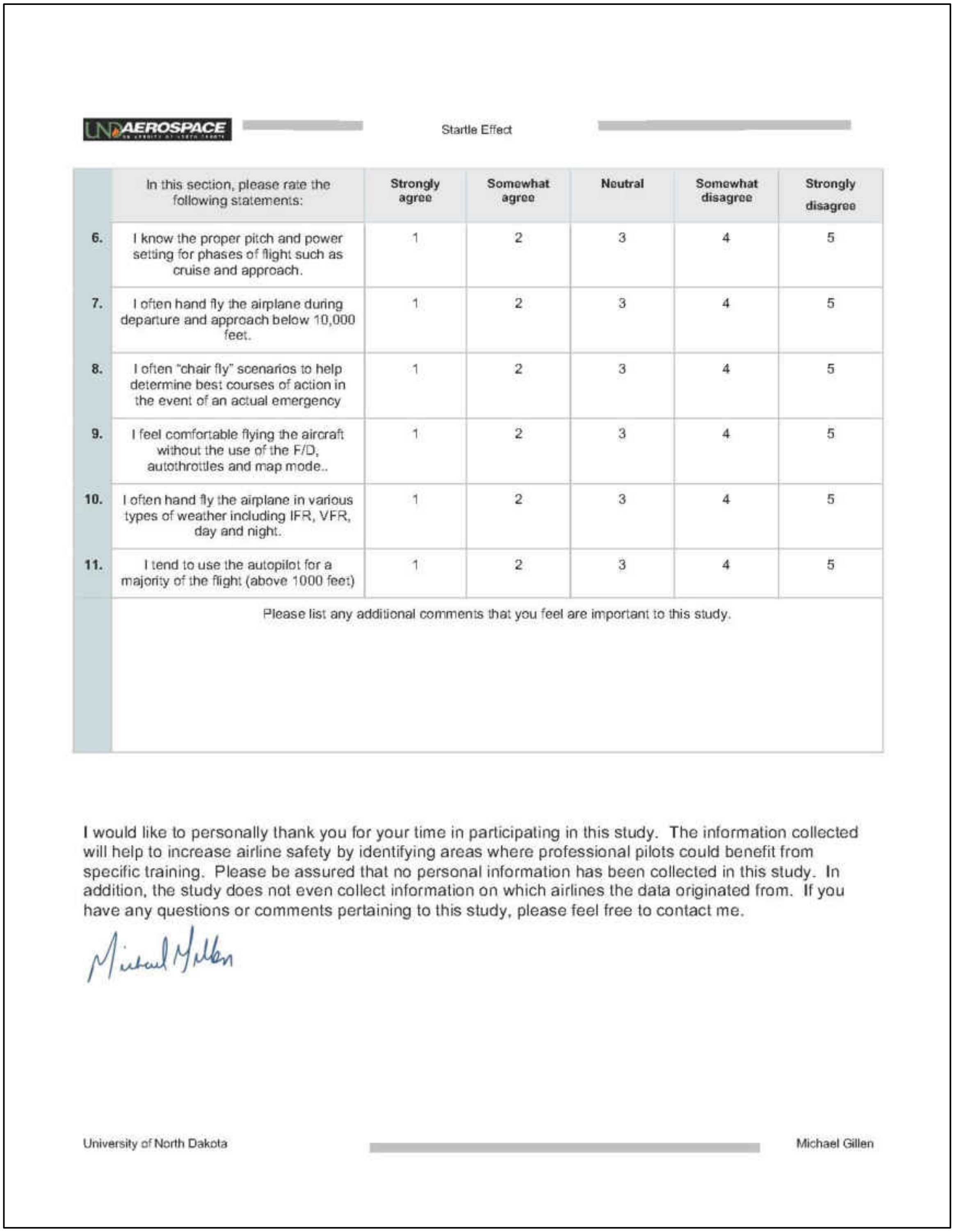
Pilot Survey

I agree that I have volunteered for this study and have felt no undue pressure from the airline, the University of North Dakota, or the principle investigator to participate. I have also been informed that no data will be kept linking any simulator performance to a specific pilot. Data collected is for this research project only and will not be reported to any entity or airline. Final aggregate results may be viewed in the published dissertation that will be available at the University of North Dakota Chester Fritz library. The researcher has informed me that I will fly as part of a crew and may encounter some usual situations in the simulator. I also understand that I have the right to refuse participation or withdraw from the study at any point without a change in relationship with United Airlines, the University of North Dakota or the research team.

Training (Y/N)	**Years on Current Aircraft.**				
/	☐ **Captain**	☐ **First Officer**			
Airplane	☐ **A319/A320**	☐ **B737**	☐ **B747**	☐ **B757/B767**	☐ **B777**
TRG CAT	☐ **LANT**	☐ **LAT**	☐ **HANT**	☐ **HAT**	**Do not fill out.**

1.	Do you fly outside of your current job? (If yes please answer the next question).	Yes	No	
2.	Did you flying in the military	Yes	No	
3.	Do you have any type of formal aerobatic training	Yes	No	
4.	Have you encountered an unusual events in flight which you would describe as being startling?	Yes	No	
5.	If you fly outside of your current job, do you fly under instrument flight rules?	Yes	No	

Michael Gillen – University of North Dakota



UND AEROSPACE

Startle Effect

	In this section, please rate the following statements:	Strongly agree	Somewhat agree	Neutral	Somewhat disagree	Strongly disagree
6.	I know the proper pitch and power setting for phases of flight such as cruise and approach.	1	2	3	4	5
7.	I often hand fly the airplane during departure and approach below 10,000 feet.	1	2	3	4	5
8.	I often "chair fly" scenarios to help determine best courses of action in the event of an actual emergency	1	2	3	4	5
9.	I feel comfortable flying the aircraft without the use of the F/D, autothrottles and map mode..	1	2	3	4	5
10.	I often hand fly the airplane in various types of weather including IFR, VFR, day and night.	1	2	3	4	5
11.	I tend to use the autopilot for a majority of the flight (above 1000 feet)	1	2	3	4	5

Please list any additional comments that you feel are important to this study.

I would like to personally thank you for your time in participating in this study. The information collected will help to increase airline safety by identifying areas where professional pilots could benefit from specific training. Please be assured that no personal information has been collected in this study. In addition, the study does not even collect information on which airlines the data originated from. If you have any questions or comments pertaining to this study, please feel free to contact me.

Michael Gillen

University of North Dakota

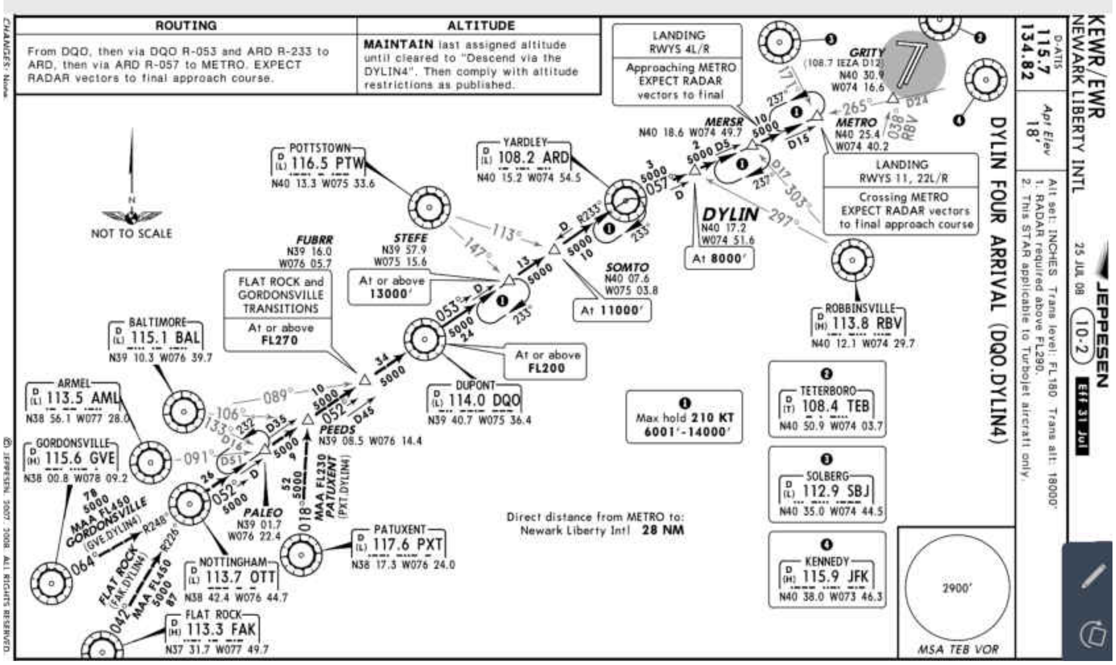
Michael Gillen

Appendix D

Simulator Setup

This section describes the simulator setup for each type of simulator used in this study. Each simulator was an FAA approved Level-D full flight simulator. The setup in each simulator was different, and the selections required to achieve the failures were also different. Crewmembers flew the simulator from their respective seat (Captain or First Officer) and were briefed to treat the simulator as they would an actual flight.

Simulator	Low Altitude	High Altitude	Notes
A320	4500 – Fuel 15 NM dogleg to final	35,000 – cruise Fail all 2 air data sources (Capt and FO).	Aircraft goes to ALT law in the high altitude scenario
B737	4300 – Fuel 15 NM dogleg to final	35,000 – cruise Fail 3 air data sources (Capt, FO, Stby)	
B747	19,000 Fuel 15 NM dogleg to final	35,000 – cruise Fail 2 air data sources	
B757	6000 – Fuel 15 NM dogleg to final	35,000 – cruise Mach/AS unreliable	
B777	12000 – Fuel 15 NM dogleg to final	35,000 – cruise Mach/AS unreliable Engine fire	High Alt – use lesson plan 14A



ROUTING	ALTITUDE
From DQO, then via DQO R-053 and ARD R-233 to ARD, then via ARD R-057 to METRO. EXPECT RADAR vectors to final approach course.	**MAINTAIN** last assigned altitude until cleared to "Descend via the DYLIN4". Then comply with altitude restrictions as published.

KEWR/EWR
NEWARK LIBERTY INTL
JEPPESEN
25 JUL 08 10-2 Eff 31 Jul
D-ATIS 115.7 134.82
Apt Elev 18'
Alt set: INCHES Trans level: FL180 Trans alt: 18000'
1. RADAR required above FL290.
2. This STAR applicable to Turbojet aircraft only.

DYLIN FOUR ARRIVAL (DQO.DYLIN4)

LANDING RWYS 4L/R
Approaching METRO EXPECT RADAR vectors to final

LANDING RWYS 11, 22L/R
Crossing METRO EXPECT RADAR vectors to final approach course

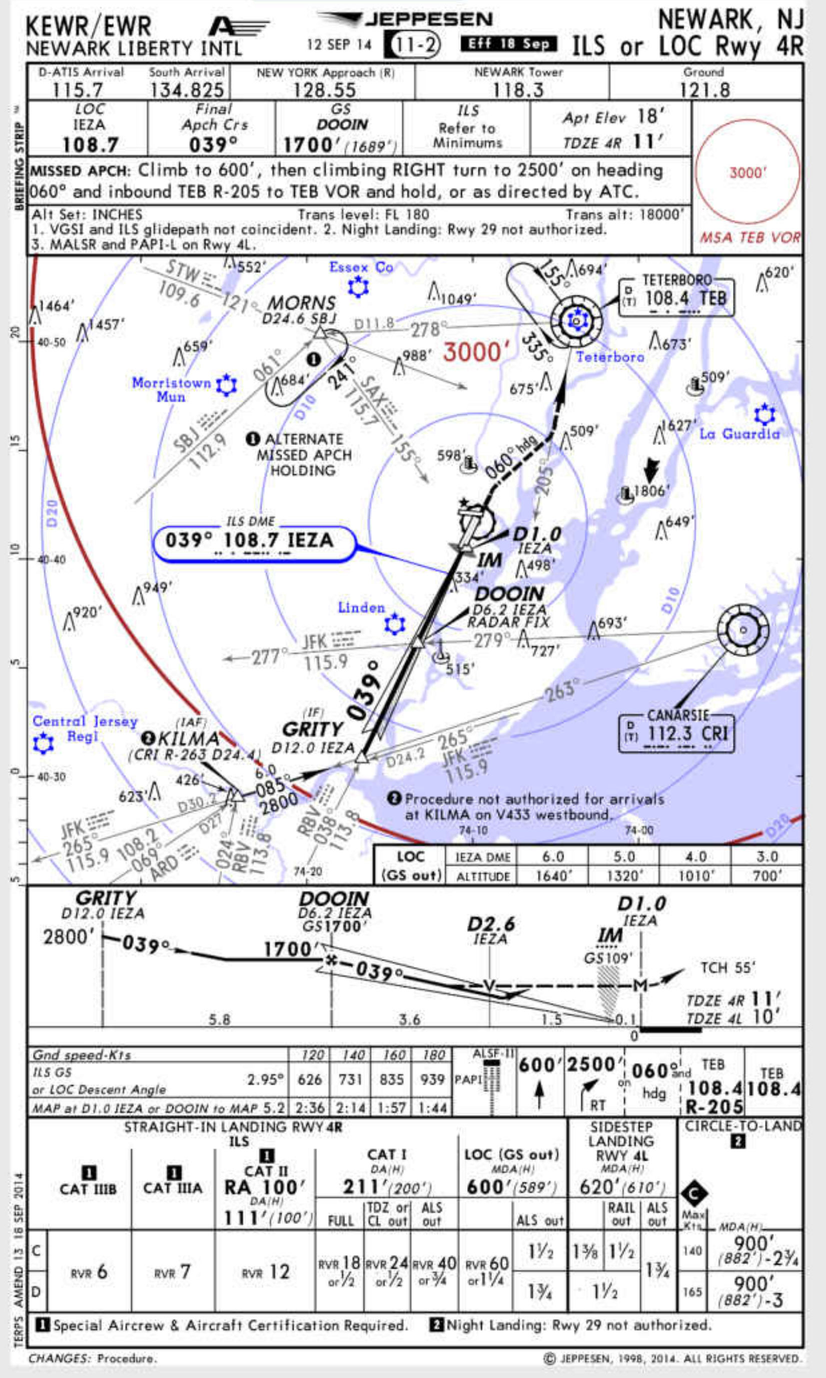
FLAT ROCK and GORDONSVILLE TRANSITIONS
At or above FL270

❶ Max hold 210 KT 6001'-14000'

Direct distance from METRO to: Newark Liberty Intl 28 NM

CHANGES: None
© JEPPESEN, 2007, 2008. ALL RIGHTS RESERVED.

DYLIN Four Arrival KEWR (Jeppesen, 2016).


KEWR/EWR
NEWARK LIBERTY INTL
JEPPESEN
12 SEP 14 (11-2) Eff 18 Sep
NEWARK, NJ
ILS or LOC Rwy 4R

D-ATIS Arrival	South Arrival	NEW YORK Approach (R)	NEWARK Tower	Ground
115.7	134.825	128.55	118.3	121.8

LOC IEZA	Final Apch Crs	GS DOOIN	ILS	
108.7	039°	1700′ (1689′)	Refer to Minimums	Apt Elev 18′ TDZE 4R 11′

MISSED APCH: Climb to 600′, then climbing RIGHT turn to 2500′ on heading 060° and inbound TEB R-205 to TEB VOR and hold, or as directed by ATC.

Alt Set: INCHES Trans level: FL 180 Trans alt: 18000′
1. VGSI and ILS glidepath not coincident. 2. Night Landing: Rwy 29 not authorized.
3. MALSR and PAPI-L on Rwy 4L.

3000′ MSA TEB VOR

LOC (GS out)	IEZA DME	6.0	5.0	4.0	3.0
	ALTITUDE	1640′	1320′	1010′	700′

Gnd speed-Kts		120	140	160	180
ILS GS or LOC Descent Angle	2.95°	626	731	835	939
MAP at D1.0 IEZA or DOOIN to MAP	5.2	2:36	2:14	1:57	1:44

ALSF-II PAPI
600′ 2500′ RT on 060° hdg and TEB 108.4 R-205 TEB 108.4

	STRAIGHT-IN LANDING RWY 4R ILS ❶ CAT IIIB	❶ CAT IIIA	❶ CAT II RA 100′ DA(H) 111′ (100′)	CAT I DA(H) 211′ (200′) FULL	TDZ or CL out	ALS out	LOC (GS out) MDA(H) 600′ (589′)	ALS out	SIDESTEP LANDING RWY 4L MDA(H) 620′ (610′)	RAIL out	ALS out	CIRCLE-TO-LAND ❷ C Max Kts	MDA(H)
C	RVR 6	RVR 7	RVR 12	RVR 18 or 1/2	RVR 24 or 1/2	RVR 40 or 3/4	RVR 60 or 1 1/4	1 1/2	1 3/8	1 1/2	1 3/4	140	900′ (882′)-2 3/4
D								1 3/4	1 1/2			165	900′ (882′)-3

❶ Special Aircrew & Aircraft Certification Required. ❷ Night Landing: Rwy 29 not authorized.

TERPS AMEND 13 18 SEP 2014

CHANGES: Procedure.

ILS 4R at KEWR (Jeppesen, 2016)

Appendix E

Additional Statistical Tests

Levene Test of Homogeneity of Variances High Altitude

High Altitude Scenario

Levene Statistic	df1	df2	Sig.
.674	1	18	.422

Regression ANOVA High Altitude Factors

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5.568	1	5.568	12.250	.003[b]
	Residual	8.182	18	.455		
	Total	13.750	19			
2	Regression	10.748	5	2.150	10.024	.000[c]
	Residual	3.002	14	.214		
	Total	13.750	19			

a. Dependent Variable: High Altitude Scenario

b. Predictors: (Constant), Crew Training Received

c. Predictors: (Constant), Crew Training Received, Problem diagnosis, Altitude control, Roll control, Pitch control

Regression Coefficients High Altitude.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	2.667	.225		11.866	.000	2.195	3.139					
	Crew Training Received	1.061	.303	.636	3.500	.003	.424	1.697	.636	.636	.636	1.000	1.000
2	(Constant)	.165	.796		.207	.839	-1.542	1.871					
	Crew Training Received	.099	.356	.059	.278	.785	-.665	.862	.636	.074	.035	.342	2.926
	Problem diagnosis	.440	.124	.591	3.561	.003	.175	.706	.812	.689	.445	.565	1.769
	Pitch control	.009	.304	.012	.031	.976	-.642	.660	.759	.008	.004	.101	9.883
	Roll control	.205	.246	.196	.834	.418	-.322	.732	.590	.218	.104	.283	3.531
	Altitude control	.186	.230	.204	.810	.432	-.307	.680	.679	.212	.101	.245	4.083

Collinearity Diagnostics[a] High Altitude

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions					
				(Constant)	Crew Training Received	Problem diagnosis	Pitch control	Roll control	Altitude control
1	1	1.742	1.000	.13	.13				
	2	.258	2.596	.87	.87				
2	1	5.598	1.000	.00	.00	.00	.00	.00	.00
	2	.304	4.290	.01	.40	.00	.00	.00	.00
	3	.049	10.693	.03	.01	.74	.00	.04	.00
	4	.029	14.012	.08	.19	.08	.04	.05	.32
	5	.016	18.762	.32	.39	.00	.16	.18	.13
	6	.004	35.529	.56	.01	.17	.80	.74	.55

a. Dependent Variable: High Altitude Scenario

Levene's Test of Equality of Error Variances[a] Low Altitude

Dependent Variable: Low Altitude Scenario

F	df1	df2	Sig.
.450	1	18	.511

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Crew Training

ANOVA[a] Low Altitude

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6.050	1	6.050	10.371	.005[b]
	Residual	10.500	18	.583		
	Total	16.550	19			
2	Regression	15.367	6	2.561	28.133	.000[c]
	Residual	1.183	13	.091		
	Total	16.550	19			

a. Dependent Variable: Low Altitude Scenario

b. Predictors: (Constant), Crew Training Received

c. Predictors: (Constant), Crew Training Received, Approach and landing, Time Management, Checklist procedures, Fuel Management, Missed approach

Regression Coefficients[a] Low Altitude

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	2.600	.242		10.765	.000	2.093	3.107					
	Crew Training Received	1.100	.342	.605	3.220	.005	.382	1.818	.605	.605	.605	1.000	1.000
2	(Constant)	.088	.371		.238	.816	-.714	.891					
	Crew Training Received	.084	.194	.046	.434	.672	-.335	.504	.605	.119	.032	.483	2.070
	Missed approach	.449	.140	.531	3.201	.007	.146	.752	.847	.664	.237	.200	5.008
	Checklist procedures	.137	.115	.145	1.186	.257	-.113	.386	.773	.312	.088	.368	2.718
	Time Magmt	.427	.118	.476	3.610	.003	.172	.683	.735	.708	.268	.316	3.165
	Fuel Mgmt	-.109	.167	-.106	-.652	.526	-.469	.252	.759	-.178	-.048	.208	4.810
	Approach and landing	.103	.124	.101	.827	.423	-.165	.370	.678	.224	.061	.366	2.736

Collinearity Diagnostics[a] Low Altitude

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions						
				(Constant)	Crew Training Received	Missed approach	Checklist procedures	Time Management	Fuel Management	Approach and landing
1	1	1.707	1.000	.15	.15					
	2	.293	2.414	.85	.85					
2	1	6.441	1.000	.00	.00	.00	.00	.00	.00	.00
	2	.367	4.189	.01	.55	.00	.00	.00	.00	.00
	3	.084	8.783	.01	.01	.11	.00	.14	.01	.06
	4	.044	12.141	.10	.02	.03	.36	.05	.07	.01
	5	.040	12.622	.35	.23	.06	.18	.10	.01	.03
	6	.016	20.048	.15	.08	.31	.00	.12	.10	.86
	7	.008	28.124	.38	.11	.48	.46	.58	.81	.04

a. Dependent Variable: Low Altitude Scenario

Levene's Test of Equality of Error Variances[a] Combined

Dependent Variable: Low and High Altitude Combined

F	df1	df2	Sig.
.046	1	38	.831

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Crew Training

Survey Correlations No Training

		Low and High Altitude Combined	I know the proper pitch and power settings.	Hand flying below 10,000 feet	Chair fly scenarios to help determine courses of action	Comfort flying raw rata	Often practice raw data skills	Autopilot usage above 1000 feet
No Crew Training	Pearson Correlation	1	.290	.101	.091	.081	.194	.373
	Sig. (2-tailed)		.228	.681	.712	.742	.425	.128
	N	19	19	19	19	19	19	18
I know the proper pitch and power settings.	Pearson Correlation	.290	1	.141	.193	.247	.022	.000
	Sig. (2-tailed)	.228		.566	.428	.309	.929	1.000
	N	19	19	19	19	19	19	18
Hand flying below 10,000 feet	Pearson Correlation	.101	.141	1	.218	.411	.534*	-.237
	Sig. (2-tailed)	.681	.566		.370	.080	.019	.343
	N	19	19	19	19	19	19	18
Chair fly scenarios to help determine courses of action	Pearson Correlation	.091	.193	.218	1	.217	.223	.248
	Sig. (2-tailed)	.712	.428	.370		.373	.359	.321
	N	19	19	19	19	19	19	18
Comfort flying raw rata	Pearson Correlation	.081	.247	.411	.217	1	.497*	.154
	Sig. (2-tailed)	.742	.309	.080	.373		.030	.543
	N	19	19	19	19	19	19	18
Often practice raw data skills	Pearson Correlation	.194	.022	.534*	.223	.497*	1	.279
	Sig. (2-tailed)	.425	.929	.019	.359	.030		.261
	N	19	19	19	19	19	19	18
Autopilot usage above 1000 feet	Pearson Correlation	.373	.000	-.237	.248	.154	.279	1
	Sig. (2-tailed)	.128	1.000	.343	.321	.543	.261	
	N	18	18	18	18	18	18	18

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

Survey Correlations with Training

		Low and High Altitude Combined	I know the proper pitch and power settings.	Hand flying below 10,000 feet	Chair fly scenarios to help determine courses of action	Comfort flying raw rata	Often practice raw data skills	Autopilot usage above 1000 feet
Crew Training	Pearson Correlation	1	-.167	-.042	-.306	-.077	.264	-.204
	Sig. (2-tailed)		.470	.857	.177	.739	.248	.375
	N	21	21	21	21	21	21	21
I know the proper pitch and power settings.	Pearson Correlation	-.167	1	.427	-.190	.283	.190	-.112
	Sig. (2-tailed)	.470		.053	.408	.214	.409	.629
	N	21	21	21	21	21	21	21
Hand flying below 10,000 feet	Pearson Correlation	-.042	.427	1	.270	.424	.694**	.016
	Sig. (2-tailed)	.857	.053		.237	.055	.000	.946
	N	21	21	21	21	21	21	21
Chair fly scenarios to help determine courses of action	Pearson Correlation	-.306	-.190	.270	1	-.085	.092	.103
	Sig. (2-tailed)	.177	.408	.237		.713	.691	.656
	N	21	21	21	21	21	21	21
Comfort flying raw rata	Pearson Correlation	-.077	.283	.424	-.085	1	.537*	.117
	Sig. (2-tailed)	.739	.214	.055	.713		.012	.615
	N	21	21	21	21	21	21	21
Often practice raw data skills	Pearson Correlation	.264	.190	.694**	.092	.537*	1	.052
	Sig. (2-tailed)	.248	.409	.000	.691	.012		.824
	N	21	21	21	21	21	21	21
Autopilot usage above 1000 feet	Pearson Correlation	-.204	-.112	.016	.103	.117	.052	1
	Sig. (2-tailed)	.375	.629	.946	.656	.615	.824	
	N	21	21	21	21	21	21	21

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

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

References

BAE. (2012). *Final Report AF447.* Mi nis tère de l ’Écologie, du Développement durable, des Transpor t s e t du Logement. Paris: Bureau d’Enquêtes et d’Analyses.

Billings, C. (1996). *Human-Centered Aviation: Principles and Guidelines.* NASA, Ames Research Center. Moffett Field: NASA Technical Memorandum 11038.

Boeing. (2012). *Statistical Summary of Commercial Jet Airplane Accidents.* Boeing Company. Seattle: Boeing.

Casner, S. M., Geven, R. W., & Willliams, K. T. (2012). The Effectiveness of Airline Pilot Training for Abnormal Events. *The Journal of the Human Factors and Ergonomics., 55*(3), 477-485.

Causse, M., Dehais, F., Peran, P., & Pastor, J. (2013, July). The effects of emotion on pilot decision-making: A neuroergonomic approach to aviation safety. . *ransportation Research Part C: Emerging Technologies, 33*, 272-281.

Coe, R. (2002). It's the Effect Size, Stupid. What effect size is and why it is important. . *Annual Conference of the British Educational Research Association, University of Exeter.* Exeter: University of Durham.

Compton, R., Banich, M., Mohanty, A., Milham, M., Herrington, J., Miller, G., . . . Heller, W. (2003). Paying attention to emotion: An FMRI investigation of cognitive and emotional tasks. *Cognitive, Affective, and Behavioral Neuroscience 3, 81-96. , 3*, 81-96.

Cummings, M. (2016). *Automation Bias in Intelligent Time Critical Decision Support Systems.* Retrieved January 12, 2016, from American Institute of Aeronautics and Astronautics: http://web.mit.edu/aeroastro/labs/halab/papers/CummingsAIAAbias.pdf

Davis, M. (1986). Pharmacological and anatomical analysis of fear conditioning using the fear potentiated startle paradigm. *Behavioural Neuroscience, 100*, 814-824.

Davis, M. (1992). The role of the amygdala in fear and anxiety. *Annual Review of Neuroscience, 15*, 353-375.

Diamond, L., & Aspinwill, L. G. (2003). Emotional regulation across the lifespan: An integrative approach emphasizing self-regulation, positive affect, and dyadic process. . *Motivation and Emotion, 27*, 125-156.

Diehl, A. (1991). The Effectiveness of Training Programs for Preventing Aircrew Error . *Sixth International Symposium on Aviation Psychology* (pp. 640-655). Columbus: The Ohio State University.

Eaton, R. (1984). *Neural mechanisms of startle behavior.* New York, NY: Plenum Press.

Endsley, M. (2006). *Situation Awareness. Handbook of Human Factors and Ergonomics.* Hoboken, NJ: John Wiley and Sons.

Ericcsson, K. (1996). *The acquisition of expert performance: An introduction to some of the issues. The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games.* Mahwah, NJ: Erlbaum .

Gillen, M. (2014). Piloting Skills Use Them or Lose Them. *Lufthansa First Annual Human Factors Conference.* Frankfurt: Lufthansa.

Hilscher, M., Breiter, E., & Kochan, J. (2012). *From the Couch to the Cockpit: Psychological Considerations During High-performance Flight Training.* Retrieved January 20, 2015, from http://apstraining.com/wp-content/uploads/Psychological-Considerations-During-High-Performance-Flight-Training-2005-Hilscher-Breiter-Kochan.pdf

Ippel, M. J. (1987). *Cognitive Task Load and Test Performance.* Retrieved February 1, 2016, from http://www.ijoa.org/imta96/paper52.html

Isaac, A. (2012). *Emergencies and Unusual Situations – Whose World View.* Retrieved January 30, 2016, from HindSight15: https://www.eurocontrol.int/sites/default/files/publication/files/120611-hs15.pdf

Jeppesen. (2016). Charts. Denver, CO.

Kaempf, G., & Klein, G. (1994). Aeronauticl Decision Making: The next generation. *Aviation Psycholoy in Practice*, 223-254.

Klein, G. (1993). *A recognition-primed decision model of rapid decision making. Decision making in action: Models and methods.* Norwood, NJ: Ablex.

Klein, G. A., Calderwood, R., & Macgregor, D. (1989). Critical Decision Method for Eliciting Knowledge. *IEEE Transactions on Systems, Man and Cybernetics, 19*(3), 462-472.

Kranse, L., & van der Schaaf, T. (2001). Recovery From Failures in Chemical Process Industry. 5(3), 199-211. *International Journal of Cognitive Ergonomics, 5*(3), 199-211.

Landis, C., & Hunt, W. (1939). *The startle pattern.* Oxford: Farrar and Rinehart.

Lang, P., Bradley, M., & Cuthbert, B. (1990). Emotion, attention, and the startle reflex. *Psychological Review, 97*, 377-395.

Layton, C., Smith, P. J., & McCoy, C. E. (1994). Design of cooperative problem-solving system for en-route flight planning: an empirical evaluation. *Human Factors, 36*, 94-119.

LeDoux, J. (1996). *The Emotional Brain.* NY, NY: Simon and Schuster.

LeDoux, J. E. (2000). Emotion circuits in the brain. . *Annual Review of Neuroscience, 23*, 155-184.

Lieberman, D. A. (2012). *Human Learning and Memory.* Cambridge, MA: University Press.

Martin, W., Murray, P. S., & Bates, P. (2012). *The Effects of Startle on Pilots During Critical Events: A Case Study Analysis.* Brisbane, Australia: Griffith University.

McKinney, E. H., & Davis, K. J. (2003). Effects of deliberate practice on decision performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 45*(3), 436-444.

McKinney, E., & Davis, K. (2003). Effects of Deliberate Practice on Crisis Decision Performance. *Human Factors, 45*, 436-444.

Merk, R. (2009). *A computational model on surprise and its effects on agent behavior in simulated environments (Technical Report NLR-TP-2009-637). Amsterdam, Netherlands: .* National Aerospace Laboratory, Technical Report NLR-TP-2009-637, Amsterdam.

MM43. (2011). Retrieved March 30, 2016, from Professional Pilot Rumor Network: http://www.pprune.org/tech-log/460625-af-447-thread-no-6-a-31.html (FDR analysis pic). Accessed March 30, 2016

Morris, C. H., & Leung, Y. (2006). Pilot Mental Workload: How Well Do Pilots Really Perform? *Ergonomics, 49*(15), 1581-1596.

Mosier, K., Palmer, E., & Degani, A. (1992). Electronic checklists: Implications for decision making. *Proceedings of the Human Factors Society 36th Annual Meeting* (pp. 7-11). Santa Monica: Human Factors Society.

Mosier, K., Skitka, L., Heers, S., & Burdick, M. (1997). Automation Bias: Decision Making and Performance in High-Tech Cockpits. *International Journal of Aviation Psychology, 8*(1), 47-63.

Muthard, E., & Wickens, C. (2002). *Factors that Mediate Flight Plan Monitoring and Errors in Plan Revision: An Examination of Planning under Automated Conditions.* University of Illinois, Institute of Aviation, Savoy. Retrieved from http://www.aviation.illinois.edu/avimain/papers/research/pub_pdfs/isap/mutwic.pdf

Neisser, U. (1976). *Cognition and Reality.* San Francisco, CO: W.H. Freeman and Co. .

Nikolic, M., & Sarter, N. (2007). Flight Deck Disturbance Management: A Simulator Study of Diagnosis and Recovery from Breakdowns in Pilot-Automation Coordination. *The Journal of the Human Factors and Ergonomics Society, 49*, 553-563.

Norman, D., & Bobrow, D. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7*, 44-64.

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

Orasanu, J., & Martin, L. (1998). Errors in Aviation Decision Making: A Factor in Accidents and Incidents. *Workshop on Human Error, Safety and System Development.* Seattle.

Parasuraman, G., & Manzey, D. (2010). Complacency and Bias in Human Use of Automation: An Attentional Integration. *Human Factors, 52*(3), 381-410.

Plant, K., & Stanton, N. (2013). What is on your mind? Using the perceptual cycle model and critical decision method to understand the decision-making process in the cockpit. *Ergonomics, 56*(8), 1232-1250.

Rasmussen, J. (1983). Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models. *IEEE Transactions on Systems, Man, and Cybernetics, 13*(3), 257-266.

Richman, H., Gobet, F., Stazewski, J., & Simon, H. (1996). *Perceptual and memory processes in acquisition of expert performance: The EPAM model. The acquisition of expert performance in the arts and sciences, sports, and games.* Mahwah, NJ: Erlbaum.

Salmon, P. M., Stanton, N., Walker, G., Baber, C., Jenkins, D., McMaster, R., & Young, M. (2008). What really is going on? Review of situation awareness models for individuals and teams. *Theoretical Issues in Ergonomics Science, 9*(4).

Sarter, N., & Woods, D. (1993). *Cognitive Engineering in Aerospace Application; Pilot Interaction with Cockpit Automation.* NASA, NASA Ames Research Center, Moffett Field.

Shappell, S., & Wiegmann, D. (2009). A Methodology for Assessing Safety Programs Targeting Human Error in Aviation. *International Journal of Aviation Psychology, 19*(3), 252-269.

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

Smith, K., & Hancock, P. (1995). Situation Awareness is Adaptive, Externally Directed Consciousness. *Human Factors, 37*(1), 137-148.

Stanton, N. A., Rafferty, L., Salmon, P., Revell, K., McMaster, R. C.-D., & Cooper-Chapman, C. (2010). Distributed Decision Making in Multihelicopter Teams: Case Study of Mission Planning and Execution From a Non-Combatant Evacuation Operation Training Scenario. *Journal of Cognitive Engineering and Decision Making, 4*(4), 328-353.

Thackray, R., & Touchstone, R. (1970). Recovery of motor performance following startle. *Perceptual Motor Skills, 30*, 279-292.

United Airlines. (2016). *Flight Operations Manual.* Denver, CO: United Airlines.

Vlasek, M. (1969). Effect of startle stimuli on performance. *Aerospace Medicine, 40*, 124-128.

Whalen, P., & Phelps, A. (2009). *The human amygdala.* New York: Guideford Press.

Wickens, C., & Flach, J. (1998). *Information processing. Human Factors in Aviation. .* (E. E.L. Wiener, & D. Nagel, Eds.) London, UK: London: Academic Press Limited.

Wiener, E. (1985). *Human factors of cockpit automation: A field study of flight crew transition.* NASA, NASA Ames Research Center. Moffett Field: NASA Ames.

Woodhead, M. (1969). Performing a visual task in the vicinity of reproduced sonic bangs. *Journal of Sound Vibration, 9*, 121-125.

Woodhead, M. M. (1959). Effect of brief noise on decision making. *Journal of The Acoustic Society of America, 31*, 1329-1331.

Yeomans, J., & Frankland, P. (1996). The acoustic startle reflex: Neurons and connections. *Brain Research Reviews, 21*, 301-314.