DATA-DRIVEN SAFETY FEEDBACK AS PART OF DEBRIEF FOR GENERAL AVIATION PILOTS

by

Nicoletta Fala

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

School of Aeronautics & Astronautics

West Lafayette, Indiana

August 2019

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Karen Marais, Chair
School of Aeronautics and Astronautics
Dr. William Crossley
School of Aeronautics and Astronautics
Dr. Steven Landry
School of Industrial Engineering and School of Aeronautics and Astronautics
Dr. Brandon Pitts
School of Industrial Engineering
Dr. Bruno Ribeiro
Department of Computer Science
Dr. Julius Keller
School of Aviation and Transportation Technology

Approved by:

Dr. Weinong Chen
Head of the Graduate Program

You start with a bag full of luck
and an empty bag of experience.
The trick is to fill the bag of experience
before you empty the bag of luck.

ACKNOWLEDGMENTS

In research, we all acknowledge that we stand on the shoulders of giants—we would not be able to advance in our body of knowledge had it not been for the researchers that have come before us. However, conducting research would be a lot more difficult if it weren't for those who help along the way in their own individual ways. I was lucky enough to have a number of mentors who have helped me become a researcher and a village that made the experience worth it.

I would first like to thank my advisor, Dr. Karen Marais, for her continued support throughout the past five years in everything I did. She has been, and continues to be, a brilliant mentor who has instilled in me a passion for teaching and research, and our conversations have greatly contributed to my intellectual growth. I also had the pleasure of having a large and multidisciplinary committee who together offered insights and advice that over time shaped my work. Their feedback, time, and support have been invaluable. I would therefore like to sincerely thank Dr. William Crossley, Dr. Steven Landry, Dr. Julius Keller, Dr. Brandon Pitts, and Dr. Bruno Ribeiro for serving on my committee and guiding my research.

My officemates, past and present, kept me sane and drove me crazy all at once. I found researchers to look up to from my first few weeks of graduate school in Payuna and Arjun. In time, and after the consumption of multiple ice cream pints, they became lifelong friends and collaborators who have not stopped supporting me through their multiple moves across the country. They have answered every phone call and talked me through all minor disasters, and I am grateful to have them in my academic and personal life. My current officemates helped get me through the stressful thesis-writing time of my life. Specifically, Divya, George, and Arpan have always managed to make life interesting. Between Harry Potter marathons, late night Age of Empires sessions, board games (Arpan will never win at Risk as far as I am concerned), weekly grocery runs, and food adventures, they have kept me on my toes at all times!

My students have taught me a great deal in class and have provided all the laughter I needed during the last year of graduate school. They have made the whole journey worth it and have solidified my aspirations for the future. The staff in the School of Aeronautics and Astronautics have been extremely helpful with everything I needed over the past nine years. Ann Broughton, in particular, adopted me into her Purdue Space Day family and has been there through all phases of my academic journey. Lisa and Morgan always took care of me and made sure I was on track with all the paperwork and ready to graduate. Anna and Kristin have been an endless source of comic relief and inspired me to go after anything I wanted, while Nicki and Rita helped turn those ideas into actions, and I am forever grateful.

I am also extremely fortunate to have a wonderful flying family that makes life fun and research relevant. Getting involved with Purdue Pilots, Inc. introduced me to lots of great pilots and made me a more confident pilot. Through the club, I got to ferry aircraft around the country and learn more about aircraft ownership than I ever thought I would need to know. Racing across the country is a bonding experience that is not easy to replicate. As a result, I am indebted to the Air Race Classic for introducing me to women of all ages who have graciously shared their stories and lessons. Cathy made for an excellent co-pilot and an even better friend—her ability to turn every negative into a positive means it’s impossible to not be in a good mood around her. Her annoying dancing and complete lack of taste buds are very refreshing, but her best quality is indubitably her completely supportive personality. Alyssa has helped my research by getting my mind off of it and making me accountable to someone else. Getting to know her has been a joy and the time we spent together made the past two years gratifying. Having friends who are directly impacted by research and improvements on aviation safety is a great motivation to work harder and I’ve had the time of my life flying alongside Mary, Alyssa, Jane, Cathy, Morgan, and Megan and I can’t wait to see them go forth and inspire new generations of pilots.

It is impossible to forget the friends I met while an undergrad at Purdue who have supported me while in graduate school. Leigh has been ridiculously proud of every tiny achievement in my grad school career, and she has managed to overcome her terrible memory to remember the important things in my life. Jeanne has stayed close no matter the distance and our long Skype calls put things in perspective and give me something to look forward to. Annie and Bryttani have demonstrated what true servant leadership looks like and have helped make me a better version of myself over the years.

Lastly, none of this would be possible had it not been for my family and all the sacrifices they have made throughout my life to give me everything I wanted to pursue. They have all endured many more drives to aerospace museum and airports than they ever thought they would have to and sleepless nights tracking airliners on Flight Aware on each of my trips across the pond to come back to school. Through my flying adventures, they have also become experts at stalking airplanes on flight radar, whether to track me down after I inevitably forgot to call or to spy on my air race competitors and pass on the information as the dedicated ground crew they have been. Thank you for learning more about aviation than any non-aviation nerd should have to!

"Flying might not be all smooth sailing, but the fun of it is worth the price." - Amelia Earhart

This research was partially funded by the US Department of Transportation/Federal Aviation Administration PEGASAS Center of Excellence under Award No 12-C-GA-PU AM44, 55. The project was managed by Michael Vu.

TABLE OF CONTENTS

ABBREVIATIONS 9
ABSTRACT 10
1. INTRODUCTION 12
1.1 Background and Motivation 12
1.2 Research and Thesis Outline 17
2. FLIGHT DATA 21
2.1 FDR Data 21
2.2 Smartphone Data 24
2.3 ADS-B Data 25
2.4 Operating Environment Information 25
2.5 FDR Data Post-Processing 26
3. DEFINING HAZARDOUS STATES AND TRIGGERS 30
3.1 State-Based Flight Representation 31
3.2 Potential Hazardous States During Takeoff 32
3.3 Characterizing Hazardous States via Measurable Parameters 35
4. PRESENTING PILOTS WITH SAFETY-DRIVEN FEEDBACK 48
4.1 Aeronautical Decision Making (ADM) 48
4.2 Flight Debrief 50
4.3 Biases in Risk Communication 53
4.4 Pilot Risk Perception Cognitive Biases 55
4.4.1 Framing Language 56
4.4.2 Representation Method 56
4.4.3 Parameter Type 57
5. EXPERIMENT DESIGN 60
5.1 Survey Design and Dissemination 60
5.2 Flights 63
5.3 Debrief Randomizer 64
5.4 Survey Questions 65
5.4.1 Perceived Risk 66

5.4.2 Motivation to Change 68
5.5 Demographics 69
6. SURVEY ANALYSIS 72
6.1 Demographics 73
6.2 Representation Method 74
6.3 Parameter Type 77
6.4 Framing Language 80
6.5 Interaction Effects 82
6.6 Proportional Odds Model 83
6.7 Qualitative Responses 84
6.8 Survey Analysis Results and Discussions 85
7. CONCLUSION AND FUTURE WORK 87
7.1 Limitations 89
7.2 Recommendations for Future Work 90
7.2.1 Hazardous States During Other Phases of Flight 90
7.2.2 Application of State Detection Algorithms to Data of Lower Resolution 90
7.2.3 Investigating How Pilots Respond to Surveys 91
7.2.4 Effect of Demographics 91
7.2.5 Using Simulator Studies in Place of the Survey 92
APPENDIX A. FDR DATA LONG COMPARISON 93
APPENDIX B. SURVEY 98
APPENDIX C. SURVEY RESPONSES 114
REFERENCES 138
VITA 143

ABBREVIATIONS

ACS	Airman Certification Standards
ADM	Aeronautical Decision Making
ADS-B	Automatic Dependent Surveillance-Broadcast
AOPA	Aircraft Owners and Pilots Association
ATP	Airline Transport Pilot
CFI	Certified Flight Instructor
DPE	Designated Pilot Examiner
EFB	Electronic Flight Bag
FAA	Federal Aviation Administration
FDR	Flight Data Recorder
FOQA	Flight Operations Quality Assurance
GA	General Aviation
GAO	Government Accountability Office
GAMA	General Aviation Manufacturers Association
IATA	International Air Transport Association
IFR	Instrument Flight Rules
IMC	Instrument Meteorological Conditions
MFD	Multi-Function Display
MVFR	Marginal Visual Flight Rules
NGAFID	National General Aviation Flight Information Database
NTSB	National Transportation Safety Board
PEGASAS	Partnership to Enhance General Aviation Safety, Accessibility, and Sustainability
PFD	Primary Flight Display
SMS	Safety Management System
VFR	Visual Flight Rules
VMC	Visual Meteorological Conditions

ABSTRACT

Author: Fala, Nicoletta. PhD
Institution: Purdue University
Degree Received: August 2019
Title: Data-Driven Safety Feedback as Part of Debrief for General Aviation Pilots.
Committee Chair: Karen Marais

General Aviation (GA) is the foundation of most flying activities and the training ground for civilian pilots, both recreational and professional. However, the safety record for GA is lacking compared to that of commercial aviation. Approximately 75% of accidents each year involve personnel factors, that is, even if the pilot was not the cause of the accident, they could have done something to either prevent it or improve the outcome.

In this research, I aim to improve GA safety through safety-driven post-flight debrief that encourages pilots to consider the risk in their flights and identify behavioral changes that could make their flying safer. Providing pilots with a debrief tool that they can use with or without a flight instructor requires that we know both what to communicate, and how to communicate it. Risk communication heuristics and biases have not been researched in the context of aviation and flight training and we therefore do not know how pilots understand or respond to debrief.

To achieve the goals of this work, I used a three-step process: (1) identify events that may put the safe outcome of a flight at risk, (2) detect those events in flight data, and (3) inform the pilot in a way that helps them improve in their future flights. I use a state-based representation of historical aviation accidents to define a list of events or behaviors that need to be communicated to the pilots, in the form of states and triggers. I use flight data to retrospectively detect these behaviors upon completion of the flight, by mapping parameters or combinations of parameters that can be calculated and tracked in the flight data to the hazardous states and triggers defined. To present these events to pilots, I created a prototype interactive debrief tool with risk information that I use in a survey to evaluate the effectiveness of feedback in different representation formats. Specifically, I evaluate the impact of three factors: representation method (graphical and

numerical), parameter type (safety and performance parameters), and framing language (risk-centric and safety-centric).

I disseminated the survey via aviation mailing lists, type groups, flying clubs, and flight training providers, end received 268 responses. The survey analysis showed that the feedback representation does affect its effectiveness in terms of risk perception, but not when it comes to pilots' motivation to change. The lessons learnt from this survey can be used in creating additional surveys that delve further into risk communication biases and our understanding of how pilots perceive risk and feedback.

1. INTRODUCTION

The goal of this work is to improve GA safety proactively by providing pilots with feedback that will guide them towards improvements in their future flights. This Chapter motivates the current research and discusses past research upon which build my work.

1.1 Background and Motivation

Aviation is a fast-growing industry, with IATA expecting that air passengers will double to 7.8 billion in 2036 (IATA, 2017). The FAA forecasts an average 1.9% U.S. carrier passenger growth over the next 20 years (FAA, 2018). The increasing demand for air travel is creating an increasing need for pilots—Boeing (2018) reports that North America alone will require 206,000 new pilots, with 790,000 new pilots required worldwide. General Aviation (GA), the foundation of most flying activities, is usually the starting point for new pilots, both recreational and professional, while they are undergoing their initial training. While the FAA forecasts that the GA turbine and rotorcraft fleets will grow, they also forecast that the fixed-wing piston GA aircraft fleet will shrink at an average annual rate of –0.8%, as a result of the increasing cost of aircraft ownership and an aging fleet. The light-sport-aircraft category, however, is forecast to grow at an annual rate of 3.6% (FAA, 2018). When it comes to the pilot population, the FAA projects that the number of active GA pilots will decrease by ~22,600 pilots, whereas the Airline Transport Pilot (ATP) category is expected to increase by the same number. Stakeholders in aviation have voiced concerns that the supply of available and qualified pilots is inadequate to support the current or future demand from U.S. airlines, both at the regional and mainline level, resulting in a need to reduce flights or eliminate routes to some markets (U.S. GAO, 2018). The urgent need for more pilots at the regional airline level is creating a deficit in the number of flight instructors available to train new pilots and therefore the time they have available to spend with students, making it difficult for the industry to meet future needs in terms of growth and safety rates.

To keep up with the demand for new pilots amidst the economic obstacles, we need to train a higher number of pilots faster, which could result in compromised safety. Even though GA safety has improved over the past years, several hundred pilots still lose their lives in GA accidents each year. In 2017, fixed-wing GA had a total of 966 accidents, 167 of them fatal (AOPA Air Safety Institute, 2018). Figure 1 and Figure 2 show accident trends for non-commercial fixed-wing GA and commercial fixed-wing GA for the past ten years. If the number of operations increases to accommodate for the higher projected demand, we can expect to see more accidents and more fatalities if the accident rate remains constant, making GA safety a pressing concern.

Risk management is a decision-making process used to identify hazards systematically, assess the degree of risk, and determine the best course of action. While everything involves risk, unnecessary risk that has no possible benefit should not be accepted (FAA, 2008). The level of risk is most often characterized in terms of severity and probability, where severity refers to the consequences of an event occurring, and probability is the likelihood of the event. Risk includes identified risks, which have been determined and can therefore be mitigated, and unidentified risks, some of which become identified if an accident or incident happens, and some of which are never known. The level of risk may vary for each pilot, depending on their experience level and certifications. For example, a flight in Marginal VFR (MVFR) conditions may be risky for a VFR pilot, since the conditions might become Instrument Meteorological Conditions (IMC) during the flight. However, for an IFR pilot, the same flight may not be as risky, since they have the training to complete the flight, even if IMC occurs.

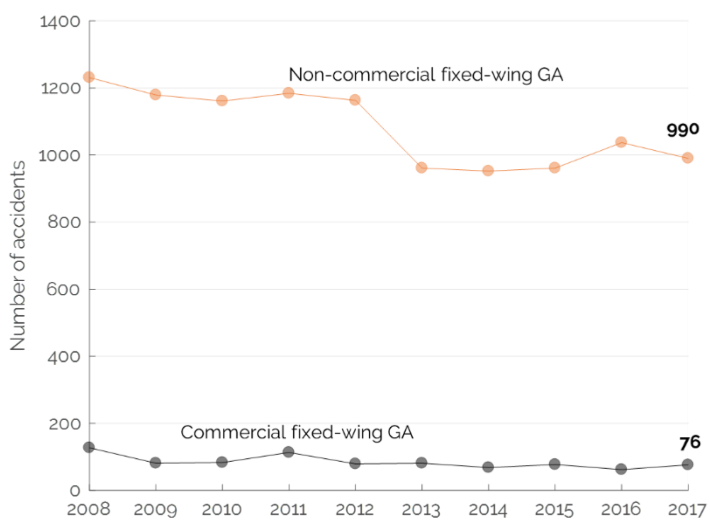


Figure 1: Non-commercial GA results in approximately ten times as many accidents as commercial GA. Although there is a large decline in 2013 for non-commercial GA, the number of accidents has remained stable since then. Adapted from (AOPA Air Safety Institute, 2018).

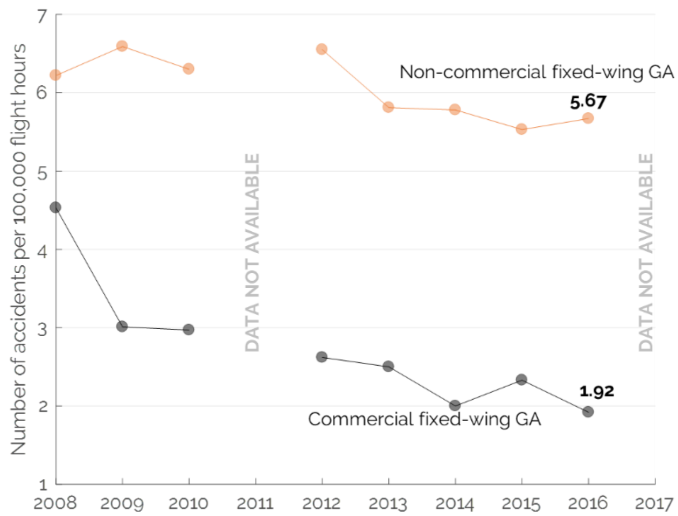


Figure 2: The FAA provides estimates for the number of hours the fleet flew under each operation category per year. We can use those numbers to calculate accident rates per 100,000 flight hours. We observe that non-commercial GA still ranks higher than commercial GA in terms of how many accidents they encounter. Adapted from (AOPA Air Safety Institute, 2018).

There are multiple approaches to improving safety in aviation, spanning aircraft and pilot certification, aircraft technology, operation, and regulation. Angle of attack indicators, for example, were designed to help prevent stalls and improve approach by providing a visual representation of the lift. Ballistic recovery systems (or parachutes) for small airplanes can be deployed to lower the aircraft to the ground, decreasing fatalities in parachute-equipped aircraft accidents (Alaziz et al., 2017). Some of the advancements to aviation safety were the result of NTSB recommendations after accidents. The NTSB has issued over 5000 aviation-related safety recommendations since its beginning (Sumwalt & Dalton, 2014). To understand the causes of accidents and incidents, the FAA uses data reactively (after the incident) while implementing a proactive approach (the Safety Management System (SMS) approach) where safety personnel analyze data and identifies and mitigates risks before they result in accidents. The FAA began implementing the SMS approach in 2005 to analyze aviation safety data and identify conditions that may lead to accidents/incidents, and mitigate the risks through changes to organization, processes, management, and culture (Dillingham, 2013). When it comes to GA, there are challenges that incumber the FAA's efforts to assess and improve safety. The GA fleet, which makes up for 90% of the U.S. civil aircraft fleet, is very diverse, with more than 220,000 aircraft in the active GA fleet. The FAA is also faced with

GA data limitations; estimates of annual GA flight hours may be unreliable (for example, there is no data provided for 2011, making it impossible to calculate accident rates for that year, as shown in Figure 2), and information on GA pilots in inadequate. The FAA is therefore unable to determine the effect of training on pilot behavior and competence, or link training to the likelihood of an accident (Dillingham, 2013).

Recent advances in flight training technology provide opportunities to rethink training operations in terms of efficiency and safety. All civilian flight students undergo their initial pilot training in a school that falls into one of three categories: collegiate aviation, non-collegiate vocational pilot school, or instructor-based pilot school. Collegiate aviation schools offer a 2- or 4-year undergraduate degree in an aviation major along with the commercial pilot certificates and ratings. Vocational pilot schools are very structured in their training sequence and curriculum. Instructor-based schools are more flexible, and the training sequence and curriculum depends on the student's needs. All pilots must meet minimum certification standards independently of the type of training they undertook. One of the key challenges of collegiate aviation schools is the recruitment and retention of flight instructors. Collegiate flight programs implement new technologies that help them align their training with airline operations (Babb, 2017). As a result, collegiate fleets tend to be equipped with Flight Data Recorders (FDRs) and SMS programs, with a streamlined electronic dispatch program and use of Electronic Flight Bags (EFB) so that the school can keep track of the flight operations. When trying to improve the efficiency and effectiveness of flight training operations, we can use the technology that pilots have in the aircraft with them.

Airlines may use Flight Operational Quality Assurance (FOQA) programs to attempt to improve both the safety and efficiency of their operations. FOQA programs analyze exceedances, which are deviations from defined expectations (Chidester, 2003). If an airplane is equipped with flight data monitoring equipment, the FOQA program will point out if any parameters exceeded boundary values (Veillette, 2014). Implementing a similar program in GA may improve safety and operational performance, maintenance procedures, and flight training (Mitchell et al., 2007). The NTSB highlighted the need to expand the use of recorders to enhance transportation safety in 2016 (NTSB, 2016). However, avionics in the GA fleet are not as advanced as those in commercial airliners. The size, weight, and cost of FDRs has precluded their use in GA in the past. Nevertheless, with improvements in technology, more and more small aircraft are equipped with FDRs and glass cockpit displays. Smartphones and tablets have made their way into the cockpit through EFB applications that help pilots with flight planning and resource management. Therefore, even in aircraft that are not fully equipped, we can still record some information about the flight. We can use the flight data to proactively improve GA safety both on an individual and community level by detecting unsafe behaviors, instead of reactively making improvements based on lessons learned from aviation accidents. Using multiple sources of data from equipment available would help address some of the challenges the FAA is facing pertaining to flight information data.

Flight schools are trying to move to a more data-centric instructional approach. Utah State University uses the avionics suite on their fleet to record and monitor flight data that they then use for training and safety. When flight data indicates that any flight exceeded limitations five or more times during a flight, the student and instructor are required to meet with the safety committee for remedial training (Utah State University, 2017). There is an opportunity for flight instructors to use the data capabilities of aircraft in post-flight debrief and objective evaluation. Commercial products that take advantage of the addition of technology in the flight decks of small aircraft to collect flight data and present pilots with a visualization of their flights, like CloudAhoy and CirrusReports, are becoming more prevalent in debrief.

At the same time, many GA pilots are flying recreationally and do not have the same resources as pilots who are still pursuing flight training or professional pilots. After they complete the necessary level of training, recreational pilots are no longer flying with an instructor. Instead, they are potentially flying as the most experienced pilot in the aircraft, and often the sole pilot. We can use flight data to continue providing pilots with debrief, like they would if they were still pursuing flight training. A good debrief "allows individuals to discuss individual and team-level performance, identify errors made, and develop a plan to improve their next performance" (Salas et al., 2008, pp. 518-527) so, by eliminating the debrief aspect of flying, we are removing the continuous learning from the flight experience. The natural order of human processing consists of experiencing something, then reflecting on it, followed by discussing the event with others, before learning from it and modifying behaviors (Fanning & Gaba, 2007). Although pilots may naturally reflect after a flight where they learnt something (satisfying the reflection aspect of debrief), it will likely not be systematic, and it may not occur at all, depending on the pilot's ability to focus.

Debrief may move through three stages: description, analysis, and application. Without a facilitator (in this case, a flight instructor) it may be hard to move on from the description phase (Fanning & Gaba, 2007). Debrief, as a learning tool, is designed as a systematic approach to reflection and discussion, and has been shown to improve performance (Tannenbaum & Cerasoli, 2013).

One potential way to improve GA safety would therefore be to continue providing pilots with debrief and feedback on their flying even beyond their flight training period, to encourage them to analyze their flights in more detail and learn from events in the flight. In the absence of a flight instructor, debrief has to be driven by flight data. Ideally, debrief tools can help the “post-instructor” pilot get a debrief like what they would be getting with an instructor. At the same time, research directed towards the development of such tools may also make flight training more efficient, by giving instructors a data-driven approach to help guide their debrief conversation. Commercial debrief products, such as CloudAhoy, focus on flight visualization and refrain from any discussion of risk, flight safety, or performance. By not discussing significant elements of the flight, such as safety, and focusing instead on visualization, they may not be pushing pilots towards the application stage of debrief. O’Hare’s Aeronautical Risk Judgment Questionnaire (ARJQ) showed that pilots displayed low levels of risk and hazard awareness, and an optimistic self-appraisal of their abilities (O'Hare, 1990), suggesting that pilots are likely to dismiss their own risk as being inconsequential. For example, if a pilot practices a maneuver on a solo flight while collecting data, and tries to visualize the maneuver after landing, using a product like CloudAhoy, they may not realize the extent of their own mistakes or the risk in their flight. To manage risk, pilots need to perceive the risk associated with a situation or hazard and decide whether they are willing to accept this amount of risk in this situation (Hunter, 2002). Safety-driven post-flight feedback may help facilitate risk management in subsequent flights, by alerting pilots to potentially hazardous situations. However, we do not know how pilots respond to debrief or whether different formats affect their response.

1.2 Research and Thesis Outline

In this research, to help address the need for safer pilots, I take a quantitative approach to evaluating whether the presentation format used in the risk communication part of debrief matters among pilots in terms of how they perceive it. Providing pilots with a debrief tool that they can use with or without a flight instructor requires that we know (1) what to communicate, and (2) how to communicate it. To achieve this goal, I used a three-step process: (1) identify events that may put the safe outcome of a flight at risk, (2) detect those events in flight data, and (3) inform the pilot in a way that helps them improve in their future flights. To evaluate how to best inform the pilot of their unsafe events, I created a survey that I disseminated among pilots. The survey allows pilots to use a prototype tool that consists of modified CloudAhoy screens and evaluate the effectiveness of debrief feedback in each case. While this research addresses a small part of the bigger problem, it provides a starting point where we can build the rest of the work required in providing pilots of different aircraft and of various skill levels debrief opportunities that may keep them safer.

Figure 3 shows the three different parts of this work with their different sub-tasks.

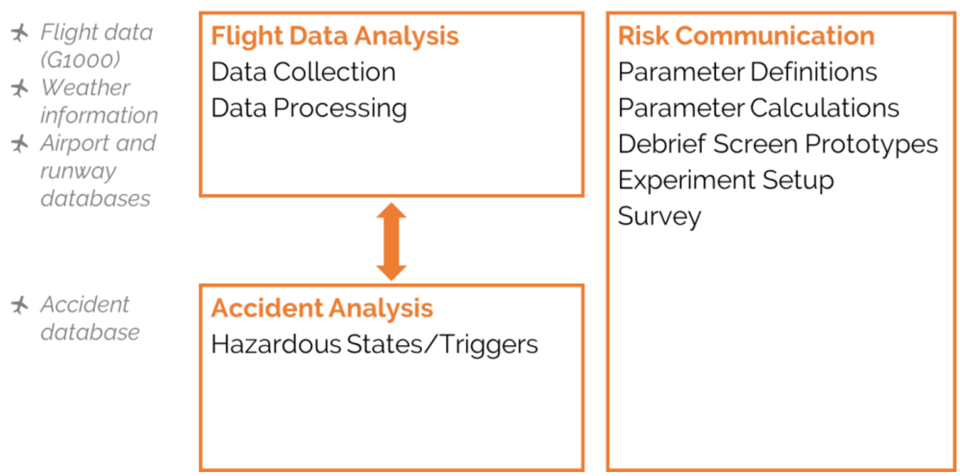


Figure 3: This research is divided in three sections. The accident analysis section identified events that tend to appear in accidents. Flight data analysis then aimed to detect these events in flight data. In this research, I mainly used Garmin G1000 flight data. The last section, risk communication, evaluated how to best communicate risk information to pilots through a debrief survey disseminated among different aviation groups.

In this thesis, structured as follows, I focus on how we can use different sources of data to make the most of what is available in GA in proactively communicating risk information to pilots.

Chapter 1 has introduced the challenges in GA safety and how they affect the research.

The purpose of Chapter 2 is to describe the flight data that I use and explain how we can counteract some of the challenges with data quality or quantity.

Chapter 3 maps introduces the hazardous state model and maps accident information, in the form of hazardous states and triggers, to events and parameters that we can calculate and detect in various forms of flight data.

I use a state-based representation of historical aviation accidents to define a list of events or behaviors that need to be communicated to the pilots, in the form of states and triggers. Each flight consists of states, nominal or hazardous, and trigger events (Rao, 2016). A state is a period of time during which the system, consisting of the aircraft and the pilot, exhibits a particular behavior, and a trigger is an event that causes the system to transition between two states. Hazardous states do not always result in accidents but preventing the hazardous states will also prevent the accident. In this research, I focused on the hazardous states that appear in accidents that occur during the *takeoff* phase of flight.

I use flight data to detect these behaviors, or events, retrospectively upon completion of the flight, by mapping parameters or combinations of parameters that can be calculated and tracked in the flight data to the hazardous states and triggers defined.

Chapters 4, 5, and 6 deal with the communication aspect of the work. Chapter 4 discusses the literature on flight debrief and cognitive biases in risk communication and introduces different debrief representations that I use to communicate information on hazardous states and triggers to pilots.

Chapter 5 goes over the work in setting up an experiment to determine whether different representation methods affect how pilots perceive the feedback in their flights. I present any detected states to pilots in the form of post-flight debrief feedback, with the goal of using the information to improve performance on subsequent attempts of the same tasks. In this chapter, I created a prototype interactive debrief tool with risk information based on CloudAhoy screens. To evaluate the effectiveness of feedback in different representation formats, I used an anonymous web-based survey where a sample of pilots self-debrief flights with safety information presented in different ways and assess the risk of the flight in each case. The survey also asked the pilots how likely they are to make changes to their flying as a result of the information they reviewed, to evaluate feedback effectiveness in terms of motivation to change unsafe behaviors. I demonstrated this approach on the hazardous states that are specific to the takeoff phase of flight. In this Chapter, I also discuss survey design decisions and their potential implications on the results.

Chapter 6 analyzes and discusses the results from a total of 268 survey responses and evaluates the effect of the different risk representations on risk perception and feedback effectiveness. I found that different presentation methods do impact risk perception. They survey asked pilots to debrief a total of three flights—the effects of the different presentation methods varied depending on the flight, suggesting that further work is needed to determine how to talk to pilots about the risk of their flights.

Chapter 7 concludes the work and summarizes the contributions of this research. It also provides suggestions for future work and highlights challenges and limitations in conducting survey work among pilots.

2. FLIGHT DATA

In commercial aviation, flight data retrieved from FDRs, or “black boxes”, is used in investigations of accidents and incidents and in FOQA programs. In GA, flight data can come from smaller FDRs (which are usually embedded in glass cockpit displays), smartphones or tablets, and ADS-B, among other devices. While FDRs do not make accidents more survivable, they can help preserve the history of a flight so that we can learn from all flights. As discussed in Chapter 1, hazardous states do not necessarily result in an accident. We can therefore find such hazardous states in successful flights and learn from them in an attempt to prevent them before they result in an accident.

In this chapter, I discuss different types of data that are available in aviation, and in GA in particular, both flight data as well as operating environment data. Flight data may consist of FDR data, smartphone data, or ADS-B data. Operating environment data includes weather, terrain, and airport information. I also present some of the data processing that can provide additional information, making the flight data source more useful.

2.1 FDR Data

Newer aircraft with glass cockpit displays usually come equipped with a Flight Data Recorder (FDR), which collects many parameters, depending on the aircraft and integrated flight deck system. Owners of older aircraft may also choose to retrofit a glass cockpit display. Among the GA fleet, the most common glass cockpit displays are manufactured by Garmin, Avidyne, and Aspen. The Garmin G1000 (Figure 4) and Avidyne Entegra (Figure 5) are integrated flight instrument systems, composed of two display units, the primary flight display (PFD) and the multi-function display (MFD), and are capable of recording flight data. Such displays come at a high cost, sometimes exceeding the value of the aircraft. Aircraft used in GA range widely in size and capabilities, and they vary in age. In 2014, the average age of all registered US GA aircraft was 36.7 years, with the average age of all piston single-engine aircraft being 44.8 years (GAMA, 2015). The more recent advancements in cockpit technology and avionics are therefore not always available in GA aircraft or to GA pilots. The aircraft used in GA come with different flight data collection capabilities, if any at all. The most simplistic or oldest GA aircraft do not come with any advanced avionics—some of them do not even have navigation or communication radios. Of all new piston aircraft delivered in 2006, 89% are equipped with glass cockpits (GAMA, 2006). However, given the aging fleet, the GA aircraft that have the ability to collect FDR data still only make up a small portion of the fleet. The majority of GA aircraft therefore are not equipped with glass cockpit displays or FDR and cannot take advantage of any safety enhancements that rely on such technology.

Figure 4: The Garmin G1000 is one of the most popular flight deck systems. The Primary Flight Display (PFD) on the left gives the pilot attitude and control information, such as altitude, heading, and airspeed (NextGA Aircraft, Inc., 2012). The Multi-Function Display (MFD) on the right can be adjusted to the pilot's liking, but tends to display secondary reference information, such as engine data, navigation charts, flight plans, and procedures (Ray, 2013).

Figure 5: The Avidyne Entegra is similar in functionality to the G1000 (Figure 4) but has a different user interface (Mindstar Aviation, 2016).

Both Garmin and Avidyne make data collection convenient—a USB flash drive (for the Avidyne Entegra) or an SD card (for the G1000) transfer the data log from the FDRs on-board the aircraft to a computer. The G1000 logs flight data in a comma separated (CSV) file, with the top rows dedicated to airframe, Garmin hardware and software information, and headers, as shown in the log excerpt in Figure 6. The data follows a tabular format at a frequency of 1Hz. Avidyne records data in a tab delimited text document, with the aircraft information in a two-column format. Avidyne also includes a legend which helps the user understand the format of the rest of the document, which is not as user-friendly as the G1000. The Avidyne parameters are grouped in nine sections: *eTimeInService*, *ePilotSettings*, *AhrsAndRateData*, *eAirData*, *eFlightDirectorData*, *ePriNavDetails*, *PriNavDisplayBlockText*, *eGpsPositionAndTimeData*, *ePistonEnginesData*, and *eTurbineEnginesData*. Each row in the data set consists of a time stamp, a data section identifier, which tells the user what records to expect in the row, and the values for the parameters belonging to that particular category. The frequency of the records in Avidyne systems varies depending on the data group.

#airframe_info	log_versic	airframe_	unit_softv	unit_softv	system_sc	system_ic	mode=NORMAL						
#yyy-mm-dd	hh:mm:ss	hh:mm	ident	degrees	degrees	ft Baro	inch	ft msl	deg C	kt	kt	fpm	deg
Lcl Date	Lcl Time	UTCOfst	AtvWpt	Latitude	Longitude	AltB	BaroA	AltMSL	OAT	IAS	GndSpd	VSpd	Pitch
7/11/2015	13:25:13	-04:00				905.8	30.2		29.8	0		-9.31	
7/11/2015	13:25:14	-04:00				905.8	30.2		29.8	0		-11.17	
7/11/2015	13:25:15	-04:00				904.8	30.2		29.5	0		-23.9	
7/11/2015	13:25:16	-04:00				904.8	30.2		29.5	0		-10.17	
7/11/2015	13:25:17	-04:00				905.8	30.2		29.5	0		0.98	
7/11/2015	13:25:18	-04:00				904.8	30.2		29.5	0		5.48	
7/11/2015	13:25:19	-04:00				905.8	30.2		29.2	0		11.78	
7/11/2015	13:25:20	-04:00				906.8	30.2		29.2	0		38.69	
7/11/2015	13:25:21	-04:00				906.8	30.2		29.2	0		17.74	
7/11/2015	13:25:22	-04:00				905.8	30.2		29.2	0		4.66	
7/11/2015	13:25:23	-04:00				905.8	30.2		29	0		-12.22	
7/11/2015	13:25:24	-04:00				904.8	30.2		29	0		-21.87	
7/11/2015	13:25:25	-04:00				905.8	30.2		29	0		-7.22	
7/11/2015	13:25:26	-04:00				905.8	30.2		29	0		2.17	
7/11/2015	13:25:27	-04:00				906.8	30.2		29	0		25.86	
7/11/2015	13:25:28	-04:00				906.8	30.2		29	0		10.66	

Figure 6: The G1000 FDR records a plethora of information in a tabular format at a frequency of 1Hz that we can use to characterize the safety of a flight. The number and type of parameters recorded depend on the interaction of the flight deck system with the aircraft.

Table 20 in the Appendix shows a comprehensive list of both G1000 and Avidyne Entegra parameters as they appear in the respective flight data logs. Not all parameters are available for both systems. For example, the cylinder head temperatures, exhaust gas temperatures, and communication frequency, are not provided for aircraft equipped with Avidyne Entegra systems, while bug settings for fundamental instruments (such as altitude and heading bugs) and angle rates, among others, are not available with the G1000. Additionally, parameters are expressed in different unit systems. For example, fuel flow is expressed in pounds per hour for Avidyne systems, but gallons per hour for G1000 systems. The Avidyne headers are also missing some of the unit specifications.

Data logs may also differ for each aircraft. Aircraft differ in their capabilities and can therefore record more or fewer parameters. The simplest example of this is the addition of another engine, which would duplicate some of the parameters. In the G1000 data set, the parameters corresponding to each engine will be prefixed by E1, E2, etc. In the Avidyne Entegra data set, engines are referred to via the suffix L or R, meaning Left or Right engine.

2.2 Smartphone Data

Smartphones and tablets have made their way into the cockpit, and pilots use them to check the weather, file flight plans, navigate, and study procedures and checklists. The same devices can be used concurrently to record data. Some applications, such as Foreflight and MyFlightBook, already include navigation data recording capabilities. Depending on the sensors available on the smartphone or tablet, these devices can record navigation information (GPS coordinates, altitude, groundspeed, ground track) and attitude information (gyroscope, accelerometer, magnetometer sensor data converted to attitude and heading information).

The data collected on a smartphone is only a subset of the data we can collect on an FDR, which affects the number of hazardous states and triggers that we can detect from it. Connecting smartphones and tablets to other portable devices, such as portable ADS-B in or portable AHRS units in combination with post-processing techniques can help provide additional data or make the current data more accurate (Chakraborty et al., 2019).

2.3 ADS-B Data

The FAA has issued a mandate requiring all aircraft operating near Class B airspace to be equipped with ADS-B Out by 2020, which can provide researchers with a plethora of flight data sets. ADS-B Out is a transponder that broadcasts aircraft parameters to ground-based towers and surrounding aircraft that are appropriately equipped. Unlike FDR and Smartphone data, the pilot does not need to provide their ADS-B data—researchers are able to collect data online or using their own receivers.

ADS-B provides position information (GPS coordinates and altitude), ground track and heading, and velocity (ground speed and vertical speed). As with the smartphone data, we can only detect a subset of the hazardous states and triggers using ADS-B data.

Even though smartphone and ADS-B data is available on more aircraft than FDR data, it is not as complete and does not provide as much information as the FDR record would. As a result, it is not possible to use the smartphone or ADS-B data to detect all the possible states and triggers; the states that may appear in a smartphone or ADS-B dataset are rather a subset of the states that would appear in an FDR dataset of the same flight. The lack of information on some states may result in pilots assuming that they do not exist, suggesting that a given flight may look safer than it actually is.

2.4 Operating Environment Information

Other data sources make it possible to expand the flight data by adding information such as weather. FDR data includes navigation information, but no information on the surrounding area. For example, FDR data does not include the name or identifier of the departure or arrival airport for a flight, or the clearance from a given obstacle. We can expand the set of hazardous states that are detectable in FDR data during post-processing to add more fields. The Airports and Runways databases provide the coordinates of each airport in the US, coordinates for the start and end of each runway, as well as information on runway lengths, widths, elevation, type (asphalt, turf, etc.), and condition, among other parameters. The FAA maintains a *Digital Obstacles File*—a database of all known obstructions within the U.S. that includes coordinates, height above the ground and above sea level, and structure information. FDR datasets include wind information at each timestamp—smartphones and ADS-B are not able to provide that information. Glass cockpit displays provide true airspeed and heading through the aircraft sensors, and groundspeed and ground track from the GPS, and then calculate wind direction and velocity from the information they have available. Smartphones, tablets, and ADS-B do not have access to the aircraft sensors and therefore cannot provide true airspeed. However, if we know the winds aloft at a particular location, we can calculate true airspeed using trigonometry.

2.5 FDR Data Post-Processing

Figure 7 shows a sequence of algorithms that post-process FDR data before it can go through flight analysis to detect any hazardous states and evaluate its risk in an automated progression.

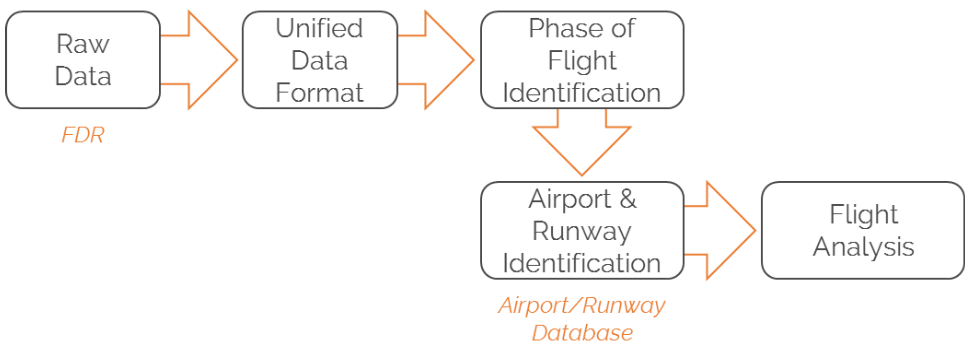


Figure 7: Independently from the source of raw data, all flight data has to go through a sequence of automated post-processing algorithms which output risk information to go into the pilot's debrief.

The first step is to unify the data format—even for data that comes from the same display, (e.g., Garmin G1000) the fields as well as their order depend on the aircraft on which the data was recorded as well as the hardware and software version on the display. Creating a unified data format is necessary to automate the algorithms that follow.

Next, I add a field to the flight data that identifies the phase of flight that the aircraft was operating under at each timestamp (Goblet et al., 2015). The phase of flight identification is important for two reasons. First, it facilitates the addition of airport and runway information for departure and arrival airports in each flight, as well as any airports visited while en route. Second, some hazardous states and triggers are relevant for particular phases of flight—deviation from the runway centerline is specific to the takeoff and landing phases of flight, whereas a low airspeed state is applicable for all phases of flight.

Table 1: The second step in data processing is to append each timestamp in the flight data with a phase of flight code, as described here (Goblet et al., 2015). There are nine unique codes corresponding to nine phases of flight that may appear in a flight.

Phase of Flight	**Phase of Flight Code**	**Definition**
Standing	2	Any time before taxi or after arrival while the aircraft is stationary.
Taxi	3	The aircraft is moving on the ground prior to takeoff and after landing.
Takeoff	4	From the application of takeoff power, through rotation and to an altitude of 35 feet above runway elevation.
Climb	5	Any time the aircraft has a positive rate of climb for an extended period of time.
Cruise	6	The time period following the initial climb during which the aircraft is in level flight.
Descent	7	Any time before approach during which the aircraft has a negative rate of climb for an extended period of time.
Approach	8	From the point of pattern entry, or 1000 feet above the runway elevation, to the beginning of the landing flare.
Landing/Touchdown	9	From the beginning of the landing flare until the aircraft touches down and exits the landing runway, or comes to a stop on the runway, or when power is applied for takeoff, depending upon the intended action after landing.
Go-around	10	A Go-around is a situation where the pilot is about to make a touchdown but decides to apply full power before the landing gear touches the ground.

Once I identify the phases of flight in each flight, I can parse the takeoffs and landings and detect the airports at which they occurred. I use the coordinates of a takeoff or landing point and designate a small bounding box around it. I then compare the dataset of airports and runways against the bounding box to see which runways the aircraft could possibly have used. Usually, only one runway happens to fall within the bounding box. Depending on the airport layout, however, if the starting points of two runways are located very close to each other, the algorithm may identify two possible runways. I compare the aircraft heading to the runway orientation to either confirm the choice of runway or choose the correct runway from the set of possible runways. In the case of an intersection departure, the algorithm may detect the wrong runway, as is the case with aircraft taking off from the intersection of Runway 27L and Taxiway D at KOSU, where (now closed) Runway 32 also happens to begin, as shown in Figure 8. To correct this misidentification, the algorithm increases the tolerance on the bounding box until it finds a runway that corresponds to the appropriate heading. Figure 8 indicates the takeoff point in an orange dot and the beginning of each runway in a blue dot. The smaller bounding box that is centered on the takeoff point has to increase in tolerance until it includes a blue dot (the bigger bounding box). Once it identifies one correct departure or arrival runway, the algorithm saves information such as the runway length and width, runway heading, runway condition, and the airport identifier.

The output of the automated algorithm progression shown in Figure 7 is an amended dataset for each flight that was processed, which includes the flight data information with fields in a specific order and with unique identifiers, the phase of flight information at each timestep, and information on the airport and runway used for each takeoff and each landing in the flight.

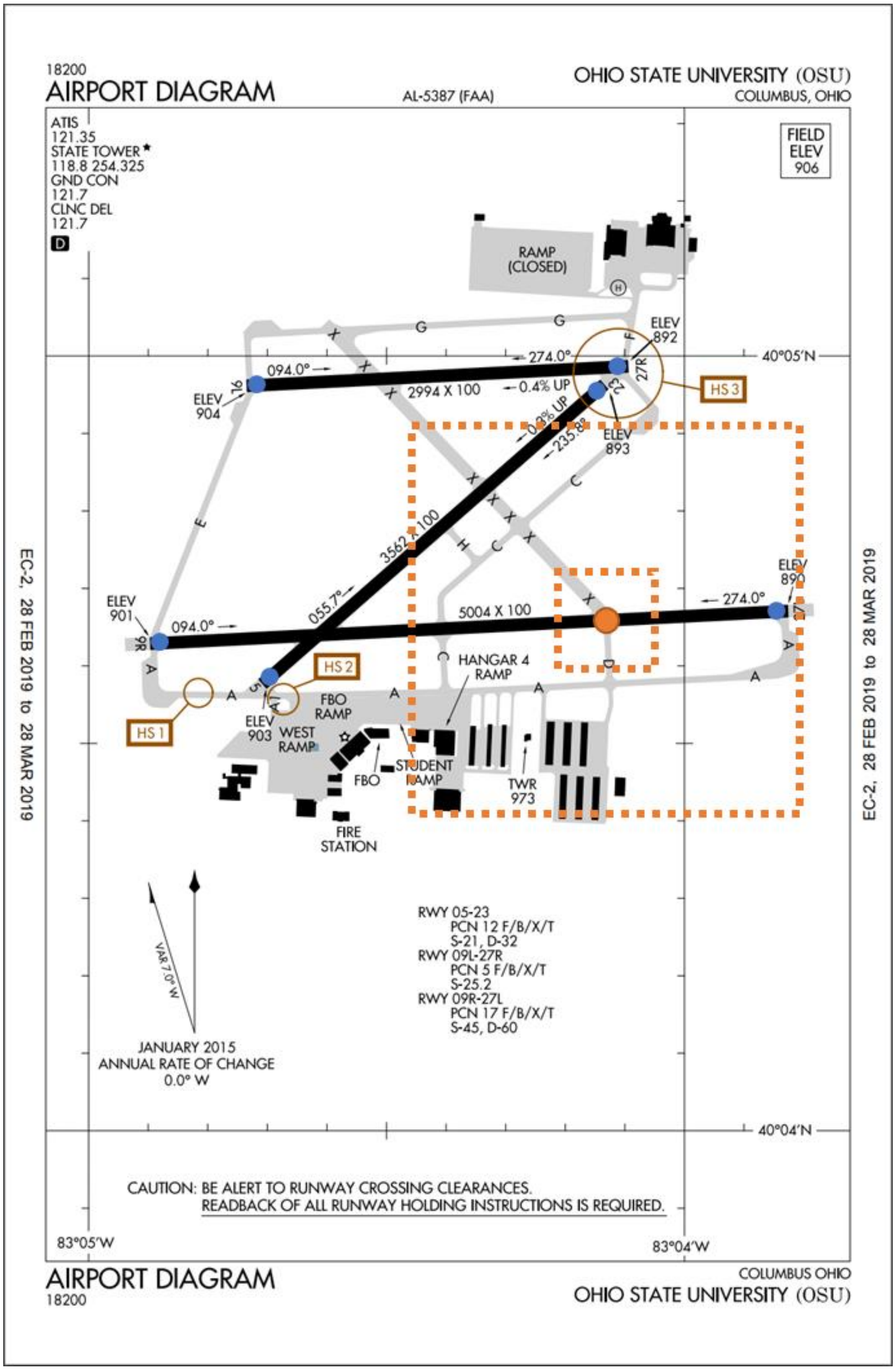


Figure 8: The Ohio State University airport's layout (FAA, 2015) creates complications in the part of the algorithm that determines from which airport and runway the aircraft took off. If the aircraft takes off from the Runway 27L and Taxiway D intersection, the algorithm outputs Runway 32 as the takeoff runway. Additional checks therefore help confirm the runway selection.

3. DEFINING HAZARDOUS STATES AND TRIGGERS

Using flight data to proactively improve GA safety requires that we are able to (1) identify events that may put the safe outcome of a flight at risk, and (2) detect those events in the flight data and inform the pilot in a way that helps them improve in their future flights. There are different ways to retrieve useful information from flight data. Detecting exceedances in the data, for example, consists of applying limits (upper or lower) on different parameters in the flight data. Safety events are off-nominal operations, or deviations from normal flying conditions, that could lead to accidents (Fala & Marais, 2016). Where exceedances only consider independent parameters, safety events combine parameters to detect various off-nominal events. For example, a bank angle of 40° is not an exceedance, and an airspeed of 70 knots is not an exceedance, but a bank angle of 40° while at an airspeed of 70 knots is a safety event (Fala & Marais, 2016). Safety events can therefore provide more safety information from flight data than exceedances, however, they still only provide information on what happened.

Anomaly detection can be used to support airline FOQA programs in the airlines by identifying anomalous flights without pre-defining parameter thresholds (Li et al., 2011). Cluster analysis algorithms are used to find patterns in datasets and detect when a particular flight differs from what has already been observed. A system that provides safety analysts with a list of flights that were tagged as anomalous together with the reasons that they were deemed anomalous can help safety experts discover human factors issues in aviation (Budalakoti et al., 2006). The variability and diversity in GA flights makes such a task more difficult; it is not always possible to have a "normal flight" pattern from which anomalous flights can differ. Anomaly detection can be useful on parts of the flight, such as the approach segment, or the pattern around different runways, where the timeseries can be normalized. As opposed to airline operations, which have pre-defined routes that aircraft follow, GA pilots operate at more airports, making routes more diverse and less populated. Under Visual Flight Rules (VFR) operations, in particular, GA pilots can choose their own form of navigation, meaning that they won't always fly in a straight line directly from airport to airport. They may choose to alter their flight plan to avoid terrain or other traffic, or to find a more scenic route. Lastly, anomaly detection in GA can identify flights that look different, which may not have a correlation with flights actually being unsafe.

3.1 State-Based Flight Representation

I model each accident or incident using a state-based representation. Each flight consists of states (nominal or hazardous), and trigger events. A state is a period of time during which the system, consisting of the aircraft and the pilot, exhibits a particular behavior, and a trigger is an event that causes the system to transition between two states (Rao, 2016).

Not all flights that involve hazardous states will result in accidents—in fact, most of them will not. Figure 9 represents the state-based model of one such flight that transitioned to a hazardous state and back to a nominal state through a remedial action. A high pitch attitude can result in a flight in the slow airspeed state, which if not corrected, can result in an aerodynamic stall. If corrected, via a remedial action trigger, such as decreased pitch attitude, the flight can return to a nominal state.

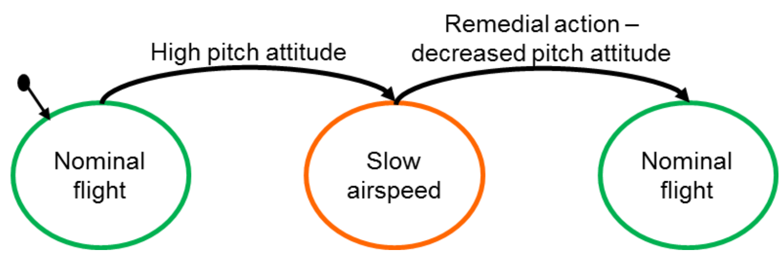


Figure 9: In this flight, the state-based flight model shows that the *slow airspeed* state was triggered by a high pitch attitude, however a remedial action returned the aircraft to the *nominal flight* state, ending with a safe landing.

In the flight modeled in Figure 10, the pilot does not take remedial action, and therefore transitions to a stall state. Inadequate recovery from the stall can result in a collision with terrain accident. Most flights do not result in accidents. They are either nominal flights, which do not enter any hazardous states, or flights that enter hazardous states but successfully recover and land safely. Hazardous states that may result in accidents also show up in flights that recovered back to the nominal state.

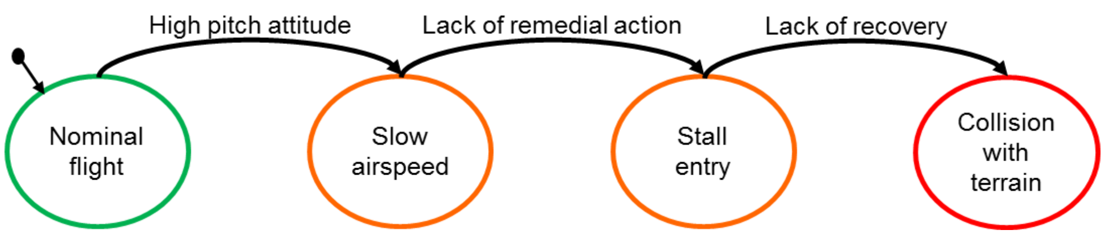


Figure 10: In another flight, the state-based flight model shows how the same *slow airspeed* state eventually transitioned to an accident through lack of remedial action and recovery.

Since accidents tend to be the result of hazardous states from which the flight never recovered, an analysis of historical accident data can help populate a list of states and triggers definitions. We can use historical accident data, as coded by the NTSB, to translate accident causes and factors into hazardous states and triggers. Another way to define additional hazardous states and triggers necessary to model flights is by using the pilot's operating handbook as well as any manufacturer recommendations. For example, manufacturers require pilots to maintain an airspeed under a *maximum airspeed* threshold for each aircraft, to avoid aerodynamic flutter. *Fast airspeed* is therefore a candidate for a hazardous state. Our flight physics knowledge can also contribute to our collection of states and triggers. *Exceeding the critical angle of attack* is a trigger event that can result in an *aerodynamic stall* state. Lastly, we can implement surveys to obtain additional events that can contribute to the list of states and triggers. These surveys can ask flight instructors and experienced pilots who are active in the GA community for states and triggers that they think may result in accidents.

Using historical accident data as a starting point for generating a list of hazardous states and triggers ensures that the considered events have the potential of being a factor in an accident, since they have already appeared in an accident. However, other sources of hazardous state definitions should not be discounted, as they allow us to detect situations that could potentially be problematic, even if they have not caused an issue yet.

3.2 Potential Hazardous States During Takeoff

In this research, I use the takeoff phase of flight to investigate the effect of representation on post-flight debrief. The takeoff phase of flight provides a good demonstration of the research because it includes a wide variety of hazardous states that capture decision making, aircraft control, and performance. To generate a list of hazardous states and triggers that are of importance during the takeoff phase, I used the FAA Airman Certification Standards (ACS) in combination with the NTSB accident database. Several pilot actions contribute to a successful takeoff, including maintaining aircraft control while on the ground and above the runway, choosing the required engine settings, lifting off at the appropriate airspeed (rotation speed), and not veering off the runway. FAA Designated Pilot Examiners (DPE) use the ACS to evaluate student pilots (FAA, 2017). The certification standards for normal takeoffs are shown in Figure 11. To obtain the hazardous states and triggers applicable to the takeoff phase, I mapped the standards from the ACS to the initial list of hazardous states and triggers from the historical accident data. For example, *Confirm takeoff power* and *proper engine and flight instrument indications prior to rotation*, is mapped to the *Insufficient takeoff power state*. Table 2 lists the hazardous states and triggers that may be present during the takeoff phase and maps them to the corresponding standards from Figure 11.

Task	***Task A. Normal Takeoff and Climb***
References	FAA-H-8083-2, FAA-H-8083-3, FAA-H-8083-23; POH/AFM
Objective	To determine that the applicant exhibits satisfactory knowledge, risk management, and skills associated with a normal takeoff, climb operations, and rejected takeoff procedures. ***Note:*** *If a crosswind condition does not exist, the applicant's knowledge of crosswind elements must be evaluated through oral testing.*
Knowledge	The applicant demonstrates understanding of:
PA.IV.A.K1	1. Takeoff distance.
PA.IV.A.K2	2. Takeoff power.
PA.IV.A.K3	3. Atmospheric conditions.
PA.IV.A.K4	4. Wind conditions and effects.
PA.IV.A.K5	5. The application of V_X or V_Y and variations with altitude.
PA.IV.A.K6	6. The manufacturer's recommended emergency procedures for relating to the takeoff sequence.
Risk Management	The applicant demonstrates the ability to identify, assess and mitigate risks, encompassing:
PA.IV.A.R1	1. Selection of runway based on wind, pilot capability, and aircraft limitations.
PA.IV.A.R2	2. The demonstrated crosswind component for the aircraft.
PA.IV.A.R3	3. Windshear.
PA.IV.A.R4	4. Tailwind.
PA.IV.A.R5	5. Wake turbulence.
PA.IV.A.R6	6. Go/no-go decision-making.
PA.IV.A.R7	7. Task management.
PA.IV.A.R8	8. Low altitude maneuvering.
PA.IV.A.R9	9. Wire strikes.
PA.IV.A.R10	10. Obstacles on the departure path.
PA.IV.A.R11	11. A rejected takeoff and predetermining takeoff abort criteria.
PA.IV.A.R12	12. Handling engine failure during takeoff and climb.
PA.IV.A.R13	13. Criticality of takeoff distance available.
PA.IV.A.R14	14. Plans for engine failure after takeoff.
PA.IV.A.R15	15. Sterile cockpit environment.
Skills	The applicant demonstrates the ability to:
PA.IV.A.S1	1. Verify ATC clearance and no aircraft is on final before crossing the hold line.
PA.IV.A.S2	2. Verify aircraft is on the assigned/correct runway.
PA.IV.A.S3	3. Ascertain wind direction with or without visible wind direction indicators.
PA.IV.A.S4	4. Determine if the crosswind component is beyond the pilot's ability or aircraft manufacturer maximum demonstrated value.
PA.IV.A.S5	5. Position the flight controls for the existing wind conditions.
PA.IV.A.S6	6. Clear the area; taxi into the takeoff position and align the airplane on the runway centerline/takeoff path.
PA.IV.A.S7	7. Confirm takeoff power; and proper engine and flight instrument indications prior to rotation:
PA.IV.A.S7a	a. Retracts the water rudders, as appropriate, confirm takeoff power and proper engine instrument indications prior to rotation, establishes and maintains the most efficient planning/lift-off attitude, and corrects for porpoising and skipping (ASES, AMES)
PA.IV.A.S8	8. Rotate and lift-off at the recommended airspeed and accelerate to V_Y (or other speed as appropriate for aircraft).
Task	***Task A. Normal Takeoff and Climb***
PA.IV.A.S9	9. Establish a pitch attitude that will maintain V_Y +10/-5 knots (or other airspeed as appropriate for aircraft).
PA.IV.A.S10	10. Retract the landing gear and flaps in accordance with manufacturer's guidance.
PA.IV.A.S11	11. Maintain takeoff power and V_Y +10/-5 knots or to a safe maneuvering altitude.
PA.IV.A.S12	12. Maintain directional control and proper wind drift correction throughout the takeoff and climb.
PA.IV.A.S13	13. Comply with responsible environmental practices, including noise abatement and published departure procedures.
PA.IV.A.S14	14. Complete the appropriate checklist.
PA.IV.A.S15	15. Comply with manufacturer's recommended emergency procedures related to the takeoff sequence.

Figure 11: Airman Certification Standards are designed to create consistent evaluation standards for pilots and examiners alike. The standards for *normal takeoffs* for a *private pilot* outline the risk management and flying skills that a private pilot candidate should be able to demonstrate when applying for their certificate (FAA, 2017).

Table 2: I identified a list of hazardous states and triggers from a subset of the NTSB database of accidents that occurred during the takeoff phase of flight. All of these states are already covered in the ACS.

Hazardous State or Trigger	ACS Mapping
Insufficient takeoff distance remaining	PA.IV.A.K1; PA.IV.A.R13
Insufficient takeoff power	PA.IV.A.K2; PA.IV.A.S7
Tailwind takeoff	PA.IV.A.K4; PA.IV.A.R4
Takeoff in high crosswind	PA.IV.A.K4; PA.IV.A.R2; PA.IV.A.S4
Deviation from centerline	PA.IV.A.S6; PA.IV.A.S12
Inappropriate runway selection	PA.IV.A.R1; PA.IV.A.S2
Inadequate airspeed at rotation	PA.IV.A.S8
High airspeed at rotation	PA.IV.A.S8
Takeoff from inappropriately short runway	PA.IV.A.K.1; PA.IV.A.R.1; PA.IV.A.R.13

3.3 Characterizing Hazardous States via Measurable Parameters

Table 2 lists the hazardous states and triggers that may appear during takeoff. Each event described in Table 2 can become a hazardous state or trigger if the associated parameters that characterize it exceed a threshold. For example, not being on the centerline becomes dangerous when the deviation is significant enough to cause a runway excursion or a collision with objects on the runway environment, such as runway lights. There are different ways of determining what the thresholds for these events should be.

The hazardous states in Table 2 can be identified either using flight data on its own, or in combination with other data. As discussed in Chapter 2, different sources of data have varying capabilities and may not be able to provide enough information for the complete set of hazardous states. Table 3 therefore shows the data required for each state, grouped according to how the parameters that characterize the state are calculated, assuming that the FDR data is available. *Simple parameters* are those that are extracted directly from raw flight data. For example, slow airspeed is a hazardous state that is characterized by a simple parameter, since airspeed is one of the parameters that is recorded by the FDR. *Derived parameters* are those that combine multiple simple parameters to make a new parameter. Pressure altitude is an example of a derived parameter, because it is based on three simple parameters: true altitude, reference pressure, and outside air temperature. *Multi-Source Parameters* may depend on weather data sources, airport information, or obstacle databases. For example, proximity to obstacle cannot be detected from flight data alone, but I can detect it by accompanying the flight data with an obstacle database, such as the FAA's Digital Obstacle File (DOF). The classification of each state may change depending on the type of flight data that is available. For example, the wind direction and velocity parameters are considered *simple parameters* when analyzing a Garmin G1000 dataset, but they could be *derived parameters* in a different kind of FDR, or even *multi-source parameters* when using smartphone data in combination with historical weather information to calculate them.

Table 3: The hazardous states and triggers during takeoff can be grouped in three parameter types (simple, derived, and multi-source parameters) based on how I calculate them.

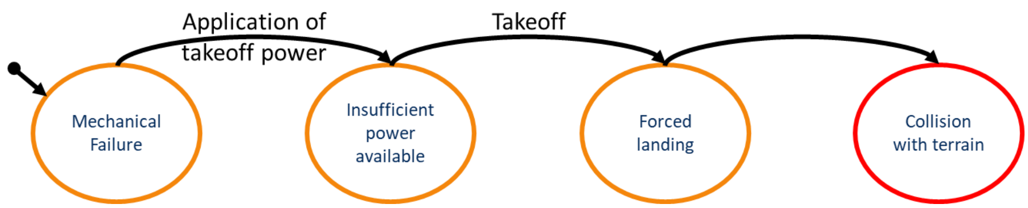
Parameter group	**State/Trigger**	**Additional data required**
Simple parameter	Insufficient takeoff power	N/A
	Inadequate airspeed at rotation	N/A
	High airspeed at rotation	N/A
Derived parameter	Tailwind takeoff	N/A
	Takeoff in high crosswind	N/A
Multi-source parameter	Insufficient takeoff distance remaining	Airport/Runway database
	Deviation from centerline	Airport/Runway database

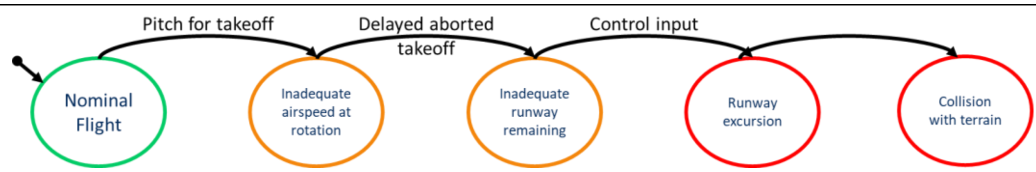
Each parameter that can be identified from data falls on a risk spectrum: sometimes a takeoff may occur in crosswinds that are slightly high (1 kt higher than recommended, for example), or in crosswinds that are much higher than recommended. Specifying *how* unsafe an event is, as opposed to just saying that it is unsafe, may help pilots understand the severity of their actions and therefore change them.

The tables in the following pages describe the hazardous states listed in Table 4, discuss their possible outcomes, and explain the process of calculating the relevant parameters so that I detect them in flight data. Note that all processes described assume that I am starting with a processed dataset as described in Section 26. The table for each state comes in two parts: state definition and state detection. *State definition* describes the motivation behind communicating the state to pilots by describing an accident where it appeared as a factor. *State detection* then discusses how to calculate parameters that can describe the state using different sources of data. Also note that the risk thresholds towards the end of the *state detection* tables are there for demonstrative purposes in this research—in flight, they can be highly dependent upon the pilot's flight training and experience and the aircraft capabilities. For example, the *Deviation from the centerline* state has risk level thresholds that are calculated based on the wingspan of a Cessna 172 and would not necessarily apply for other types of aircraft.

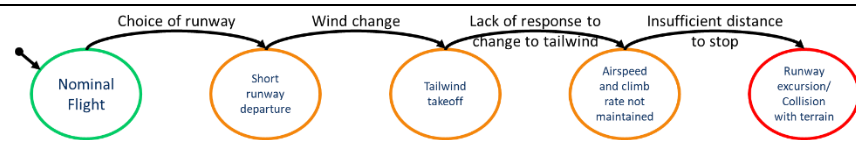
Table 4: The states presented here are adapted from Table 2 to exclude those that we cannot calculate from the flight data that is available right now. Each state has an associated table that describes how it is defined and calculated in more detail.

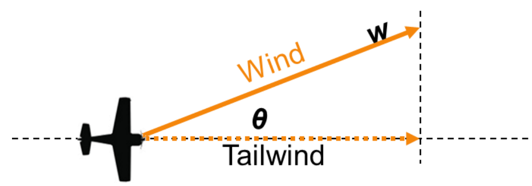
State	**Pages**	**Notes**
Insufficient takeoff power	42	
Inadequate/High airspeed at rotation	43	
Tailwind takeoff	44-45	These two states will be presented in unison in Chapters 3, 4, and 5.
Takeoff in high crosswind	46-47	
Insufficient runway distance remaining at takeoff	48-49	This parameter aims to address the *Takeoff from inappropriately short runway* state and the PA.IV.A.K.1 and PA.IV.A.R.13 ACS standards.
Deviation from centerline	50-51	

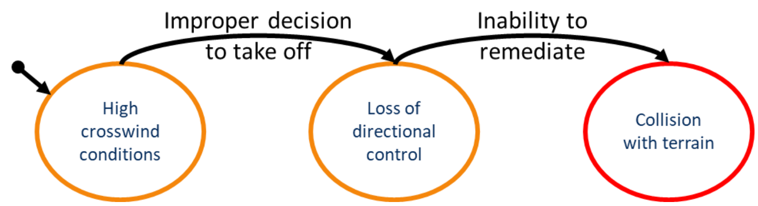
State Definition: Insufficient takeoff power	
Severity	Taking off with insufficient power limits the lift generated, resulting in an aircraft that may be unable to climb fast enough to clear obstacles. Insufficient power may be the result of mechanical issues, or the pilot having incorrect throttle/power settings. In carbureted engines, applying carburetor heat will also decrease performance.
Accident Example (**SEA88LA191**)	A flight in a Piper PA-28 resulted in an accident in Troutdale, OR, in 1988, after the pilot took off with insufficient power. The aircraft, unable to adequately climb, ended up in trees in a raspberry field, resulting in four injuries. The investigation revealed that the partial loss of engine power was a result of a mechanical failure in the exhaust carburetor system.
Accident State-Based Model	
State Detection: Insufficient takeoff power	
Parameter	*Engine RPM*
Type	Simple parameter
Data required	• Engine RPM
Risk levels	Risk level 1: *Engine RPM* < 2300 Risk level 2: *Engine RPM* < 2250 Risk level 3: *Engine RPM* < 2200

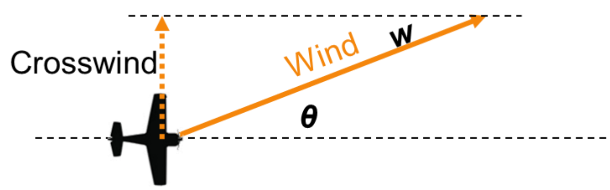
<table>
<tr><th colspan="2">State Definition: Inadequate/High airspeed at rotation</th></tr>
<tr><td>Severity</td><td>Inadequate airspeed at rotation may result in the aircraft getting out of ground effect prematurely and coming back down to the runway instead of climbing. Waiting until airspeed is too high to rotate can result in a late takeoff, with a decreased margin of safety, and an increased difficulty to maintain directional control. Increased airspeed also increases the amount of runway required to abort takeoff if needed.</td></tr>
<tr><td>Accident Example (WPR14LA250)</td><td>During the takeoff roll for a local flight in Alturas, CA, a Cessna 172RG became airborne momentarily without reaching rotation speed, and came back down to the runway. The pilot noticed that the takeoff roll was taking longer than usual, and decided to abort takeoff since the runway remaining was not enough to continue. The pilot reduced power, and both the pilot and the passenger applied the brakes, intentionally veering off the right side of the runway. The aircraft collided with a ditch and fence and nosed over. The NTSB attributed the accident to the aircraft's inability to attain rotation speed. The pilot's delayed decision to abort the takeoff also contributed to the accident.</td></tr>
<tr><td>Accident State-Based Model</td><td>


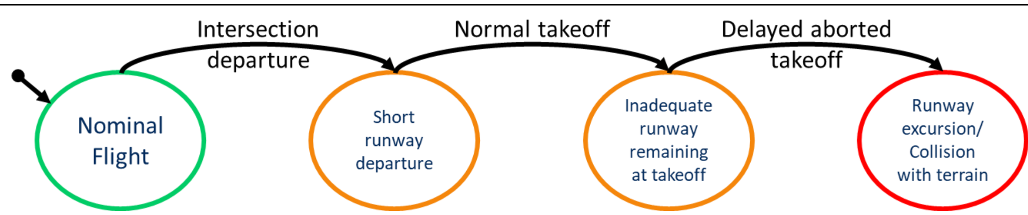
</td></tr>
<tr><th colspan="2">State Detection: Inadequate/High airspeed at rotation</th></tr>
<tr><td>Parameter</td><td>Airspeed</td></tr>
<tr><td>Type</td><td>Simple parameter</td></tr>
<tr><td>Data required</td><td>• Indicated Airspeed</td></tr>
<tr><td>Risk levels</td><td>Risk level 1: < 54 or > 56
Risk level 2: < 50 or > 60
Risk level 3: < 46 or > 64</td></tr>
</table>

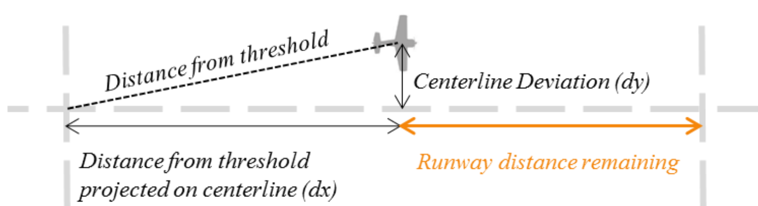
<table>
<tr><th colspan="2">State Definition: Tailwind takeoff</th></tr>
<tr><td>Severity</td><td>Taking off with a tailwind results in an increased groundspeed, which increases takeoff distance, and leads to inadequate runway remaining at rotation. The increased groundspeed and delayed rotation may result in the aircraft being unable to climb at a speed that ensures obstacle clearance.</td></tr>
<tr><td>Accident Example (CEN14LA406)</td><td>A Cessna Cardinal (177RG) collided with a tree line after an accidental tailwind takeoff in Manistee, MI, in 2014, resulting in four minor injuries. The flight instructor on board the aircraft reported that the weather station was inoperative during the preflight check, and used the airport's windsock to select the appropriate runway for takeoff, which indicated a light and variable wind, primarily from the east. They decided to use runway 19, as their intended destination was towards that direction. They also used the short-field takeoff procedure for the takeoff, and took off with approximately 1,000 ft of runway remaining. However, during the initial climb, the airplane lost airspeed and began to sink back towards the ground, touching down at the runway departure threshold and continuing into the tree line. After the accident, the flight instructor noted that the airport's windsock indicated a north-northwest wind direction with wind gusts of 18-20 knots, resulting in a tailwind condition. The corresponding decrease in airspeed and the reduced climb gradient resulted in the aircraft being unable to continue the takeoff.</td></tr>
<tr><td>Accident State-Based Model</td><td>
</td></tr>
</table>

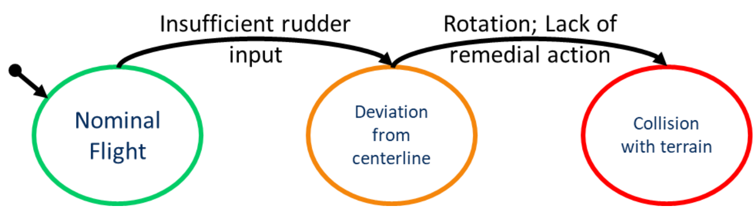
State Detection: Tailwind takeoff	
Parameter	*Tailwind at takeoff*
Type	Derived parameter
Data required	• Takeoff time • Wind velocity • Wind direction • Aircraft heading
Calculation	1. Identify the takeoff point in flight data 2. Find the corresponding wind direction and wind speed. 3. Calculate the tailwind component. $$Tailwind = w\cos\theta$$ 
Risk levels	Risk level 1: *Tailwind component* > 0 Risk level 2: *Tailwind component* > 3 Risk level 3: *Tailwind component* > 5

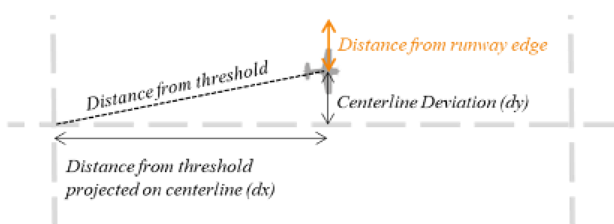
State Definition: Takeoff in high crosswind	
Severity	Crosswind during landing and takeoff can require a lot of drift correction in low airspeed conditions, where the control surfaces are less effective. Different airplanes (and different pilots) have different capabilities to counteract the crosswind drift.
Accident Example (**GAA16CA227**)	When the pilot of an American Champion Scout, a tailwheel aircraft, attempted to take off in 18 knots of crosswind at Plainview, TX, the right wing dropped shortly after takeoff, followed by a drop in the left wing, which impacted the ground. The wind gusts exceeded the aircraft's crosswind correction capabilities, resulting in the loss of directional control.
Accident State-Based Model	

State Detection: Takeoff in high crosswind	
Parameter	*Crosswind at takeoff*
Type	Derived parameter
Data required	• Takeoff time • Wind velocity • Wind direction • Aircraft heading
Calculation	1. Identify the takeoff point in flight data 2. Find the corresponding wind direction and wind speed. 3. Calculate the crosswind component. $Crosswind = w \sin\theta$ 
Risk levels	Risk level 1: *Crosswind component* > 10 kts Risk level 2: *Crosswind component* > 15 kts Risk level 3: *Crosswind component* > 18 kts

State Definition: Insufficient runway distance remaining at takeoff	
Severity	Taking off with insufficient runway remaining may lead to multiple problems: • If there are obstacles at the end of the runway, and the pilot uses up the entire runway to take off, the aircraft may not have enough time/distance to climb at an altitude that clears the obstacles at the end of the runway. • After a takeoff late down the runway, the pilot is left with less options should any mechanical problems occur. For example, if an aircraft takes off at the beginning of a runway, and the engine fails shortly after takeoff, the pilot could potentially land straight ahead on the remainder of the runway. However, if the aircraft takes off towards the end of the runway, the only option is to now find somewhere to land ahead of the runway, while also not having a lot of altitude to lose.
Accident Example (**ERA12LA314**)	In 2012, a student pilot flying a Piper Warrior, decided to do an intersection departure at Lake Wales Municipal Airport, knowing that he only needed 800 ft of runway to take off. He applied full power and let the aircraft accelerate to a rotation speed of 63 knots before pulling on the control yoke to rotate. Seeing that the airplane was not rotating, the student decided to abort the takeoff. The CFI observing the student reported that the airplane became airborne for a few seconds only 3-4 ft above the runway. Unable to bring the airplane to a stop on the runway, the student pilot veered to the right and collided with bushes in a runway excursion. At the intersection, the student pilot had 1,000 ft of runway available.
Accident State-Based Model	

State Detection: Insufficient runway distance remaining at takeoff	
Parameter	*Runway distance remaining*
Type	Multi-source parameter • Flight data (FDR, Smartphone, or ADSB) • National Transportation Atlas' Airport Runway Database (RITA, 2016)
Data required	From flight data: • GPS Coordinates at takeoff point From runway database: • GPS coordinates at the runway threshold on both ends of the runway • Runway length
Calculation	1. Identify the takeoff point in flight data 2. Detect the airport and runway from which the aircraft took off by finding airports/runways that fall in a boundary box around the takeoff point 3. Use GPS coordinates to calculate the distance between the takeoff point and the runway threshold 4. Use the coordinates of the threshold of the runway at the two ends to find the centerline 5. Project the distance from the threshold onto the centerline and correct for the Earth's curvature (*dx*) 6. Subtract the projected distance from the total runway length to obtain the *Runway distance remaining* 
Risk levels	Risk level 1: *Runway distance remaining* < 2000 ft Risk level 2: *Runway distance remaining* < 1500 ft Risk level 3: *Runway distance remaining* < 1000 ft

State Definition: Deviation from centerline	
Severity	Deviation from the runway is usually the result of insufficient rudder control while accelerating. As the pilot advances the throttle to full power, the left-turning tendencies of a single-engine airplane increase, requiring right rudder application to counteract them. Maintaining directional control on the runway is important both during the takeoff and landing phases.
Accident Example (**GAA16CA284**)	A pilot flying a Citabria, a tailwheel airplane, in Ferndale, MT, in 2016, drifted left of the runway centerline during his takeoff roll. He attempted to correct by applying right rudder, which resulted in the airplane slowing down, suggesting that the pilot was touching the brakes. The pilot released the right rudder to adjust his foot so that it would not touch the brake, and noticed that the airplane was quickly approaching the left edge of the runway. He decided to rotate early, but the airplane continued deviating towards the left, and ended up colliding with a hangar and catching fire. The NTSB reported that the cause of the accident was "the pilot's loss of directional control during takeoff, resulting in a decision to rotate early, and a collision with a hangar and subsequent fire."
Accident State-Based Model	

<table>
<tr><th colspan="2">State Detection: Deviation from centerline</th></tr>
<tr><td>Parameter</td><td>Distance from centerline/Distance from runway edge</td></tr>
<tr><td>Type</td><td>Multi-source parameter
• Flight data (FDR, Smartphone, or ADSB)
• National Transportation Atlas' Airport Runway Database (RITA, 2016)</td></tr>
<tr><td>Data required</td><td>From flight data:
• GPS Coordinates at takeoff point
From runway database:
• GPS coordinates at the runway threshold on both ends of the runway
• Runway width</td></tr>
<tr><td>Calculation</td><td>1. Identify the takeoff point in flight data
2. Detect the airport and runway from which the aircraft took off by finding airports/runways that fall in a boundary box around the takeoff point
3. Use GPS coordinates to calculate the distance between the takeoff point and the runway threshold
4. Use the coordinates of the threshold of the runway at the two ends to find the centerline
5. Project the distance from the threshold onto a line perpendicular to the centerline and correct for the Earth's curvature (dy) to obtain the Distance from centerline
6. Subtract the projected distance from the runway width to obtain the Distance from runway edge

</td></tr>
<tr><td>Risk levels (Cessna 172)</td><td>Risk level 1: $\textit{Centerline Deviation} > 0.75\left(\frac{Runway\ Width}{2} - 18\right)$
Risk level 2: $\textit{Centerline Deviation} > 0.5\left(\frac{Runway\ Width}{2} - 18\right)$
Risk level 3: $\textit{Centerline Deviation} > 0.25\left(\frac{Runway\ Width}{2} - 18\right)$</td></tr>
</table>

4. PRESENTING PILOTS WITH SAFETY-DRIVEN FEEDBACK

The third part of the research, as shown in Figure 3, deals with communicating hazardous states and triggers to pilots. Pilots are subject to cognitive biases that may affect their perception of risk and their behavior. This Chapter reviews the literature on decision making in aviation and how cognitive biases may impact post-flight debrief. I then introduce three factors which I investigate in this work in terms of their impact on feedback effectiveness.

4.1 Aeronautical Decision Making (ADM)

Approximately 75% of GA accidents involve some kind of pilot error, suggesting that the pilot could have done something to avoid or stop the accident (AOPA Air Safety Institute, 2018). Aeronautical Decision Making (ADM) provides pilots with a structured and systematic approach to analyzing in-flight changes (FAA, 1991). ADM is defined as the ability to search for and establish the relevance of all available information regarding a flying situation, to specify alternative courses of action, and to determine the potential outcomes from each alternative course of action (Jensen et al., 1987). Jensen (1995) defines *pilot* judgment as “the mental process that we use in making decisions.” The terms *judgment*, *decision making*, and *aeronautical decision making* are used interchangeably in aviation human factors research (Hunter, 2003). Decision making is one of the most important factors in human performance in aviation (O'Hare, 2003) and decisional errors are one of the major causal factors of fatal accidents (Jensen & Benel, 1977; Shappell & Wiegmann, 1997).

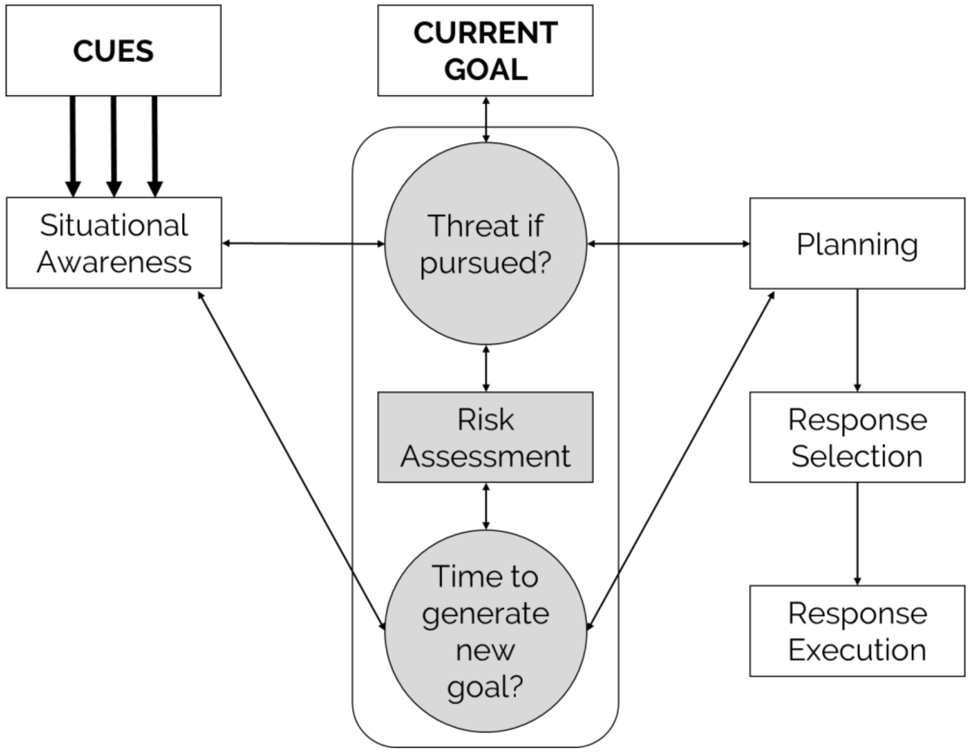


Figure 12: O'Hare (2003) identified six components of decision making in his ARTFUL decision making model. The central process of risk assessment links situational awareness and planning. The risk associated with the current goal at the top is continually assessed until the risk becomes unacceptable, at which time the decision maker can come up with a new goal if time permits.

Jensen and O’Hare have both studied and modeled aeronautical decision making and pilot judgement. O’Hare’s ARTFUL decision-making model suggests that a current goal will only be altered if the pilot’s situational awareness indicates a need for change, and there is time to generate a new goal, as shown in Figure 12 (O'Hare, 1992; O'Hare, 2003). Jensen’s Pilot Judgement Model is broken in two parts: *rational judgment* and *motivational judgment* (Jensen, 1995). Rational judgment is “the ability to discover and establish the relevance of all available information relating to problems of flight, to diagnose these problems, to specify alternative courses of action and to assess the risk associated with each alternative”, and motivational judgment is “the motivation to choose and execute a suitable course of action within the available time frame” (Jensen, 1995). Both O’Hare and Jensen have indicated that to help pilots improve their decision making, judgment, and flying habits, pilots should know and understand what they need to change, and why they should change it. For example, if the pilot is high on the approach to a runway, they first need to realize that their approach is not good enough and identify what specifically is going wrong. They should then come up with a way to fix it, such as decreasing power, or pitching down more, depending on the flight variables, and then follow through with the plan to correct the approach. If they are too close to the runway, the approach may not be salvageable, in which case the “new goal” could be a go-around.

To facilitate goal setting and decision making in aviation, flight instructors debrief the flight lesson after the flight, as discussed in next section.

4.2 Flight Debrief

During flight training, flight instructors use feedback, either during flight, or in a post-flight debrief, to communicate ways to improve performance or correct mistakes to their students. However, after a successful checkride, the now-licensed pilot no longer has an instructor or examiner by their side to talk to about their flight performance or safety, and they may not be aware (or want to acknowledge) that their actions during the flight could have resulted in an accident or incident. After they complete their initial training, some GA pilots continue their training towards more advanced certificates, while others continue flying recreationally, receiving only the minimum mandated training once every two years, which means they do not have easy access to organized feedback on their performance.

Commercial products that take advantage of the addition of technology in the flight decks of small aircraft to collect flight data and present pilots with a visualization of their flights, like CloudAhoy and CirrusReports, are becoming more prevalent in debrief. Such products can integrate flight data with other aviation resources, such as a sectional chart or instrument approach procedure plate, to display information that helps the pilot visualize their flight after they land. This type of analysis is non-evaluative, in that it only displays an objective replay of the flight in different settings, with no commentary on flight performance. For example, Figure 13 and Figure 14 show screenshots of the debrief page of a flight, as recorded using a smartphone. Figure 13 gives the pilot an overview of the flight, and the pilot can then choose what specific part of the flight they want to debrief in Figure 14 (for example, the takeoff segment).

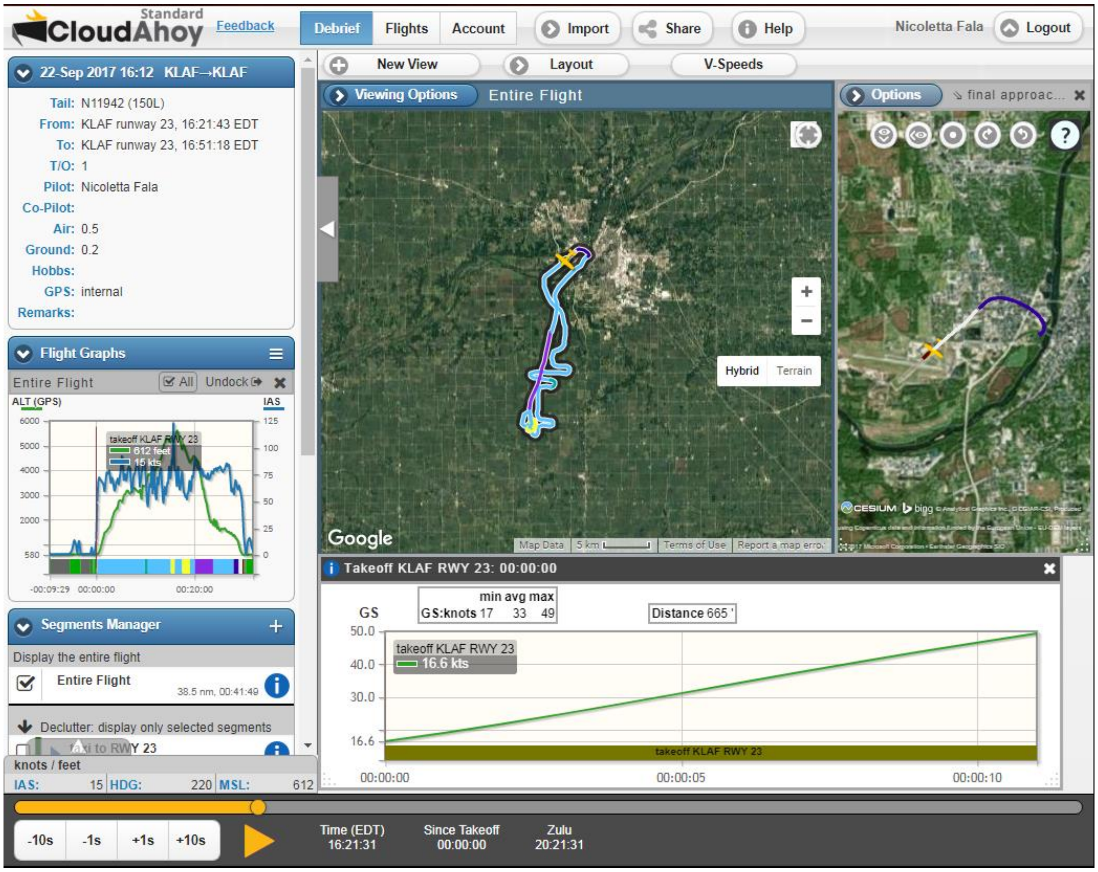


Figure 13: The Debrief screen on CloudAhoy presents pilots with an overview of their flight and allows them to choose how they want to view their flight in various windows.

In Figure 14, the pilot can see the ground track and ground speed profile of their takeoff. However, there is no indication of the quality or safety of the takeoff. While products like CloudAhoy can be helpful in reviewing flights, both during and after formal training, they do not and are not meant to provide safety guidance. A good debrief, however, "allows individuals to discuss individual and team-level performance, identify errors made, and develop a plan to improve their next performance" (Salas et al., 2008, pp. 518-527), so, by eliminating the debrief aspect of flying, we are removing the continuous learning from the flight experience.

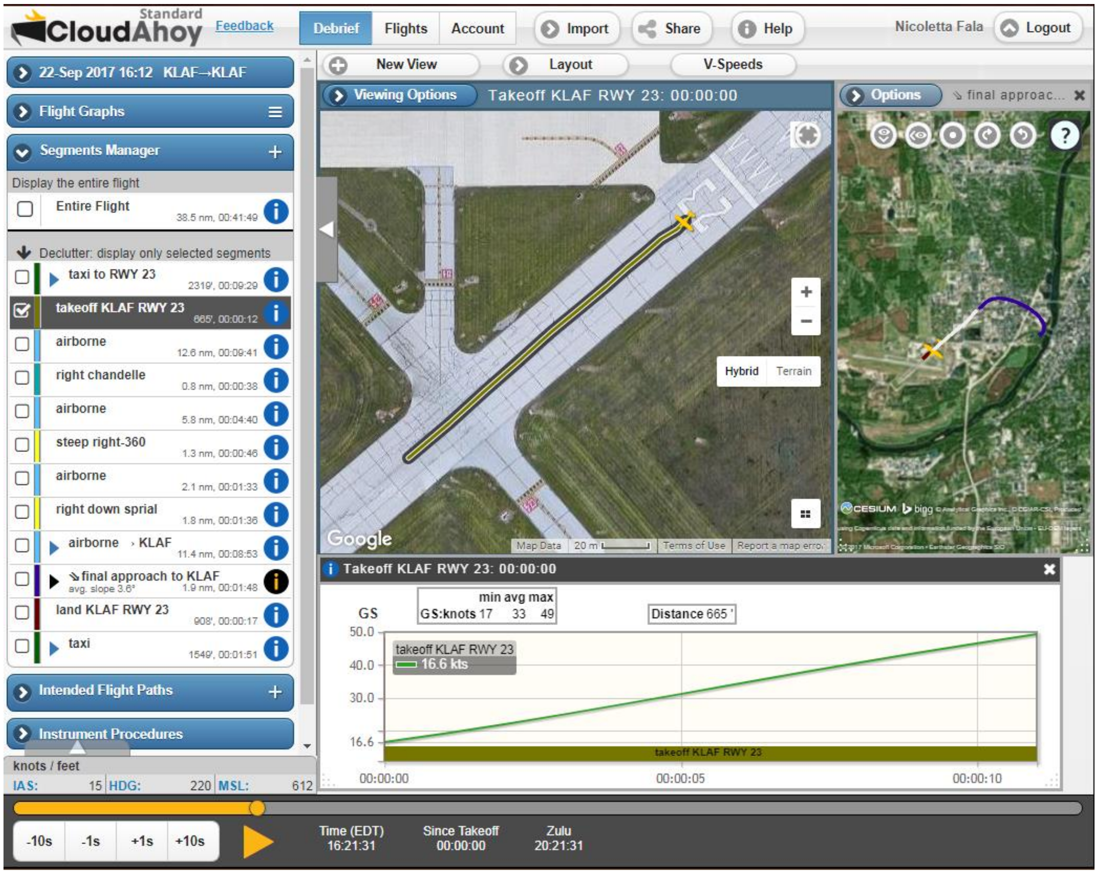


Figure 14: The CloudAhoy debrief algorithm separates the flight into segments so that the pilot can look at different flight portions individually, and tailors the information on the screens to that particular segment of flight.

Various efforts are now underway to provide safety guidance. For example, MITRE's GA Recording Device (GAARD) app records flight data from smartphone or tablet sensors to create a database of GA flight data (MITRE, 2014). The National General Aviation Flight Information Database (NGAFID) then allows the user to upload data, either from a Garmin G1000 FDR, or from the GAARD app, from a Cessna 172S or Cessna 182 airplane, and identify potential safety risks (NGAFID, 2017). While such products, services, and initiatives are in place to help pilots improve their flying, pilots who exhibit the anti-authority or invulnerability hazardous attitudes may dismiss them, or choose to justify their actions (FAA, 2016). For example, a pilot may dismiss feedback after observing that taking a particular unsafe action, such as taking off at a high airspeed, or cruising with a richer mixture than recommended, has not resulted in an accident or incident in their case. A researcher or regulator's attempt to correct such behaviors in pilots that exhibit hazardous attitudes might therefore appear alarmist to such pilots. However, as discussed in

Chapter 3, hazardous states appear in flights that did not necessarily result in accidents, and a close escape on a particular occasion does not guarantee that such escapes are always possible.

To manage risk, pilots need to perceive the risk associated with a situation or hazard, and decide whether they are willing to accept this amount of risk in this situation (Hunter, 2002). Safety-driven post-flight feedback may help facilitate risk management in subsequent flights, by alerting pilots to potentially hazardous situations.

However, there has not been enough research on effective ways to debrief flights, or on how to communicate risk information to pilots. O'Hare's (2003) work points out that "the effects of the Tversky and Kahneman (1974) work on decision heuristics and biases have been highly significant in a number of fields. Surprisingly, this has not been the case in research on aeronautical decision making." It is still not clear how pilots respond to debrief or how different formats within the debrief affect pilot response in terms of risk. In Section 4.3 I discuss cognitive biases from different fields that may be significant in aviation.

4.3 Biases in Risk Communication

Researchers in the fields of medicine, education, and sports, have studied biases to which humans are susceptible when given feedback. Physicians use different risk communication methods in attempts to convince their patients to change specific behaviors that could be hazardous to their health. Research in medicine aims to evaluate different formats of conveying health risks to patients. The intent of health risk messages is to increase perceived risk and motivate behavior change (Lipkus, 2007), similarly to how communicating risk to pilots aims to increase their understanding of risk and motivation to fly more safely. Coaches and teachers provide students with feedback so that they can improve their performance, while being careful not to hinder their progress.

The language used in feedback messages may lead the recipient of the message towards particular conclusions and bias their understanding (National Research Council, 1989). Phrases that are different, but logically equivalent, can cause individuals to change their preferences. For example, a study of how patients consent to medical procedures by Gurm and Litaker (2000) showed that framing the risk involved in a medical procedure impacted the patient’s likelihood to consent to the procedure~~.~~ Regression toward the mean, the phenomenon describing that variables that are extreme on a first measurement tend to be closer to the average on the second measurement, may also play a role in how pilots perceive the feedback. Pilots who do well on a task are likely to do more poorly on a second attempt of the same task, regardless of whether the feedback they receive is positive or negative (Kahneman & Tversky, 1973). Flight instructors are conditioned to think that they are punished for rewarding their students and rewarded for punishing them, when they do not apply regression to the mean in their reasoning. The second attempt at a flight maneuver after making a severe mistake will likely be closer to the mean.

Medical researchers have evaluated how numeric, verbal, and visual communication formats affect how likely patients are to change their behaviors. Numeric formats report the numbers of people affected by a behavior, or the probability of an event, verbal formats describe how the person involved is affected by a behavior, and visual formats present the numbers in graphs and diagrams. While the intent of each message is to communicate risk accurately and motivate behavior change, different communication formats may affect how pilots respond to feedback.

Providing feedback recipients with a lot of information may result in information overload (National Research Council, 1989). As a result, people tend to desire simplicity and therefore prefer feedback to be categorized into distinct and polar groups, rather than following a continuous scale. This categorization fosters a demand for convincing proof in feedback, suggesting that telling people that something is unsafe is not sufficient. Feedback effectiveness also depends on how it treats uncertainty and whether it bases decisions on sound science or a “better safe than sorry” attitude.

Numbers are often used in conjunction with statistical metrics to describe risk. Numeric formats appeal to people because they convey precision and accuracy. Numbers also tend to be perceived as more scientifically credible and can be verified for accuracy. People with low numeracy can have trouble understanding numerical metrics (Lipkus, 2007).

Probabilistic information can be presented in different formats: probabilities, odds, percentages, and natural frequencies. Lipkus highlighted recommendations for using numeric formats. People need a reference point that facilitates their understanding of risk. The reference point, according to Lipkus, can come in two forms: the risk of the flight resulting in an accident had the detected hazardous state or trigger not been present in the flight, and a comparison to the likelihood of a different event happening (such as the likelihood of being in a car accident). Numeric formats should be consistent: percentages should be compared to percentages, and odds to odds, and the denominator in each case should be the same (Lipkus, 2007). While comparing 5 out of 25 and 10 out of 100 may be easy for some users, other users who put emphasis on the denominator may perceive the risk differently (Paling, 2003). Small numbers approaching zero may be regarded as insignificant, and rounded numbers are more readily understood.

4.4 Pilot Risk Perception Cognitive Biases

The biases found in research in medicine and education may also be applicable to the GA pilot population. For example, assuming that pilot actions can be either safe or unsafe, to various degrees, we can describe a flight on two scales: based on how safe it was, or how unsafe it was. A flight that ranks high on the safety scale will rank low on the risk scale. While mathematically both scales are describing the same thing, pilots may perceive them differently. Risk compensation may result in pilots not taking any risk reduction measures after a flight that ranks high on the safety spectrum, whereas presenting the same flight as ranking low on a risk scale may motivate pilots to reduce their risk. At the same time, pilots may also classify a flight that is low on the risk scale (for example, 10% risky, or 90% safe) as a safe flight and dismiss the hazardous states that were present in the flight, since it was an overall safe flight. However, the actions pilots are motivated to take following feedback may differ depending on whether the feedback was framed on a safety scale or a risk scale (Fala & Marais, 2019a).

Based on the literature review on risk communication research in other disciplines, I focus this research on the three factors shown in Table 5, which may affect how pilots perceive their safety-driven feedback.

Table 5: In this research, I investigate three factors that may bias pilots in their perception of safety-driven feedback, based on a review of the risk communication literature.

Factor	Method description
Language	Risk-centric language and risk scales
	Safety-centric language and safety scales
Representation method	Graphical representations
	Numerical representations
Parameter type	Parameters that refer to the system's safety
	Parameters that refer to the system's performance

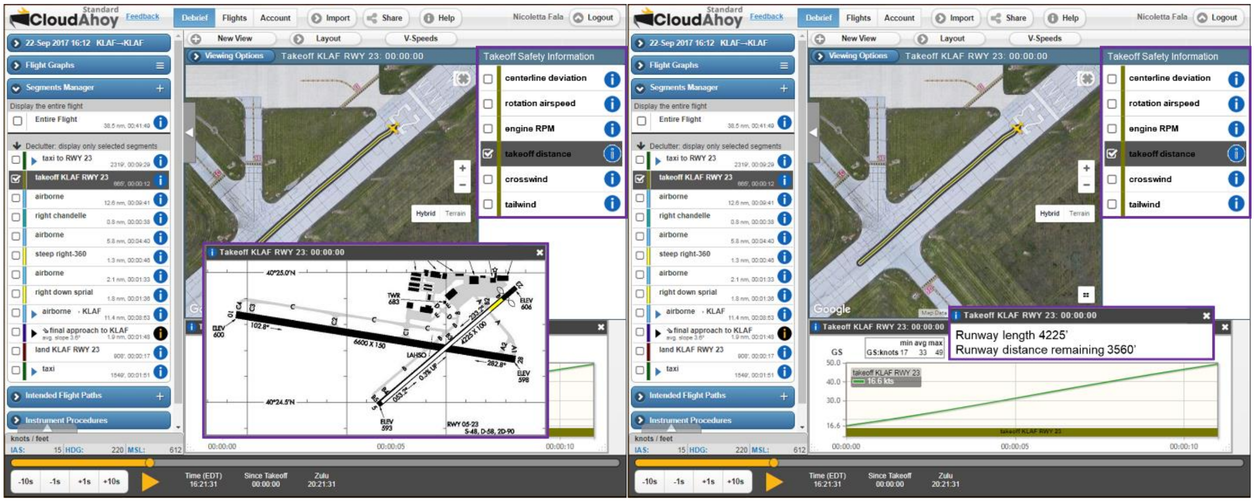
The remainder of this Chapter discusses how these three factors may affect risk perception and feedback effectiveness and how they apply to the different takeoff phase of flight states from Chapter 3.

4.4.1 Framing Language

Using language to frame a flight's risk may affect how pilots respond to their debrief feedback. Objectively, grading a flight on a risk scale (a 5-point Likert scale where 5 is extremely risky, for example) or a safety scale (with 5 being extremely safe) does not make a difference. Ranking a flight as a 4 on a 0 to 5 safety scale is mathematically the same as ranking the same flight as a 1 on a 0 to 5 risk scale. However, if the research on framing applies to the pilot population, pilots may respond more urgently to their feedback if they think of it in terms of risk, or if I present it to them using risk-centric language.

4.4.2 Representation Method

In medicine, doctors are cautious about using numerical methods to communicate risk to patients, as those methods rely upon the patient having adequate numeracy. However, the pilot population may be different than the patient population, since getting a pilot's license requires them to take a written test that includes mathematical calculations. Pilots may also prefer the exactness of quantifiable measurements as opposed to the potential vagueness of graphical representations. Figure 15 applies a graphical and numerical representation method to the *Inadequate runway distance remaining* hazardous state as an example.



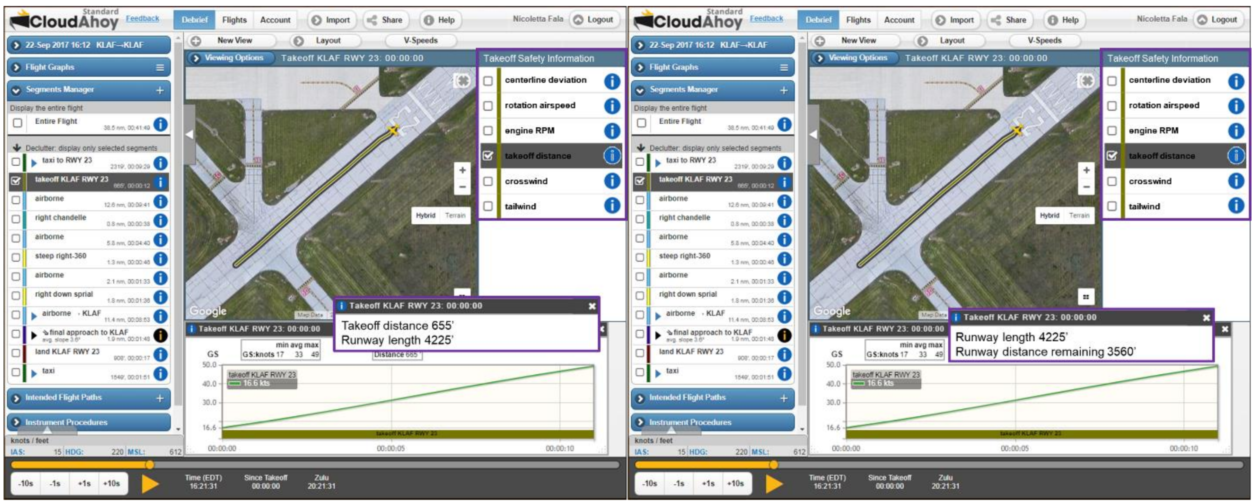
(a) (b)

Figure 15: Screen (a) uses a graphical representation to show the user where on the runway the aircraft took off, whereas screen (b) uses numbers to tell the user how much runway they had remaining after the takeoff point.

All states can be represented numerically or graphically. For example, *Engine RPM* can be stated as a peak value during the takeoff phase or plotted on a graph. *Airspeed at rotation* may be indicated on a picture of the airspeed indicator, and the *wind components* can be shown using arrows or wind sock pictures.

4.4.3 Parameter Type

Giving risk feedback to pilots on behaviors that may never result in an accident can result in them questioning and eventually ignoring the feedback. Some pilot behaviors may not be entirely risky in nature, but preventing them will still improve flying technique. For example, touching down further down a long runway is arguably not unsafe, if there is sufficient runway remaining to stop (and in some cases, it may even be preferable). However, on a shorter runway, landing further down the runway instead of closer to the beginning may result in a runway excursion if the aircraft does not have enough space to come to a stop. A pilot can still improve by landing closer to the beginning of the runway. Calling the second case a performance concern instead of a safety concern clarifies the intent of the feedback before the pilot demands proof that landing further down the runway is unsafe. Wording in behaviors that are not impacting flight safety with high certainty may therefore affect pilots' willingness to make changes to their flying. For example, landing 2,000 ft from the beginning of the runway is unsafe on a short runway but not unsafe on a long runway. Touching down 1,000 ft from the end of the runway, however, is arguably unsafe in both cases. Figure 16 displays information about the *Insufficient runway remaining distance* hazardous state framed in terms of safety and performance parameters. Measuring the distance remaining from the end of the runway informs the pilot how close they are to an unsafe situation, whereas measuring the distance actually used to take off tells them how close they were to the runway distance they calculated during their preflight based on the conditions.



(a) (b)

Figure 16: The performance parameter in the popup message in Screen (a) gives the pilot the actual takeoff distance, which is the performance parameter. Screen (b) uses a safety parameter to give the pilot the runway distance remaining at the takeoff point.

Similarly, *centerline deviation* is a performance parameter since it measures the distance from the aircraft's longitudinal axis to the runway centerline, whereas its complement, the distance measured from the runway edge, as shown in Figure 17, is a safety parameter.

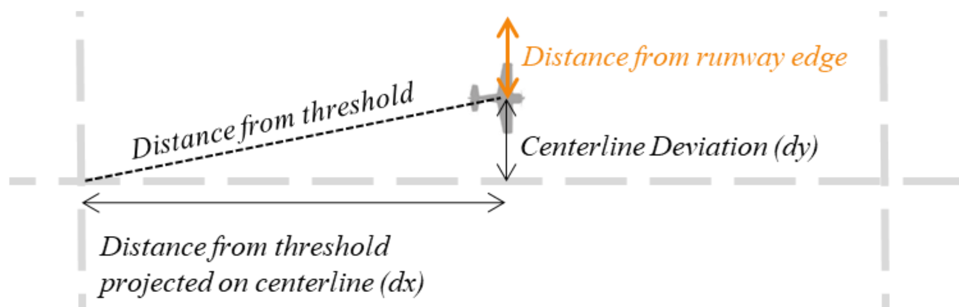


Figure 17: Lateral deviation on the runway can be measured from two reference points: in the safety parameter case, it is a measure of distance from the runway edge to the aircraft. In the performance parameter case, the deviation is the distance from the runway centerline, commonly referred to as centerline deviation.

5. EXPERIMENT DESIGN

To evaluate whether pilots are susceptible to the biases described in Chapter 3, I designed a survey that I disseminated to pilots using various aviation mailing lists and groups. The survey asked pilots to self-debrief a set of three sample flights using modified CloudAhoy screenshots and assess each flight's risk. This Chapter discusses the survey questions and dissemination, as well as the full-factorial experiment design that I used to evaluate the effect of each of the three factors described in Section 4.4.

5.1 Survey Design and Dissemination

The survey is web-based to maximize the number and diversity of potential respondents. An internet survey has the potential to collect data from a large and diverse sample of participants (Leong & Austin, 2006). A web survey gives access to individuals in distant locations or participants who may be otherwise difficult to reach (Wright, 2005). At the same time, though, web surveys also introduce biases. Self-selection bias results in a systematic bias, where some individuals are more likely than others to complete the survey, while others will tend to ignore the invitation to participate in the online survey. In my case, it is possible that self-selection bias will result in people who have a safety-mindset being more likely to respond to the survey. Nonresponse bias arises when the responses of individuals who take the survey differ from those of individuals who opt out. Such sampling issues inhibit our ability to generalize and estimate population parameters. However, the higher response rate of web-based surveys makes them less vulnerable to biases due to unrepresentative samples. In my case, a representative sample would consist of approximately 10% women, 40% private pilot license holders, and 25% commercial pilot license holders.

The survey was disseminated via various aviation groups, newsletters, and mailing lists. The Curt Lewis and Associates Flight Safety Information newsletter is distributed daily to more than 36,000 subscribers and is tailored to people with an interest in aviation safety. The Partnership to Enhance General Aviation Safety, Accessibility, and Sustainability (PEGASAS) flight schools forwarded the survey to their students. Multiple social-media based groups of pilots, such as the FAA's

General Aviation Safety group also responded to the survey. I encouraged snowball sampling by generating a constant survey link that respondents could forward to other pilots. Snowball sampling resulted in the survey being forwarded to flying clubs and the Ninety Nines. Overall, about 1,100 people accessed the survey introduction by clicking on the link.

The survey consists of three main parts, as depicted in Figure 18. Each survey starts with an *introduction* and a *tutorial*, which shows the pilots how to use the debrief tool and explains the purpose of the survey. The survey ends with the *Demographics* section, which asks pilots demographic questions, to help identify whether pilots are biased differently depending on their characteristics and experiences. The *Debrief randomizer* segment of the survey, further explained in the following sections, assigns each respondent specific feedback representation methods in flights to debrief. All pilots received the same flights to debrief, but the representation method for each flight was randomized. Respondents were able to stop taking the survey at any point.

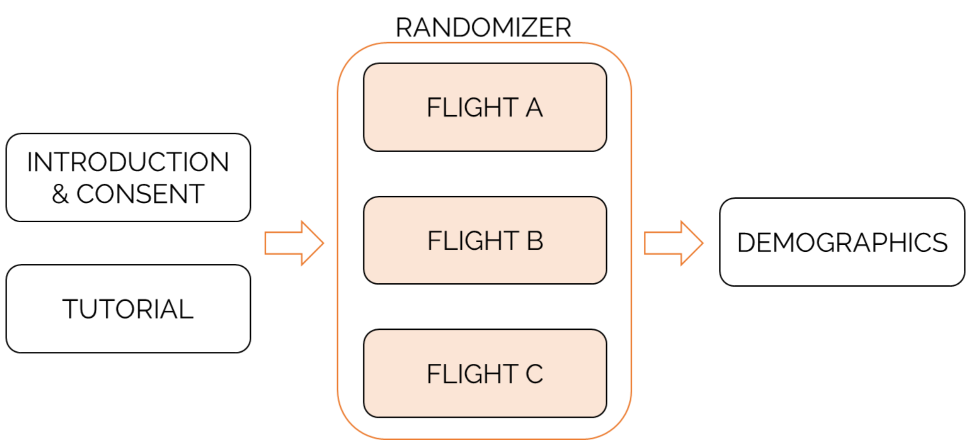


Figure 18: The white blocks in the survey structure represent the parts of the survey that are the same for every respondent. The flight randomizer in the middle allows me to evaluate whether pilots have biases by showing them data in different formats in the flights they are evaluating.

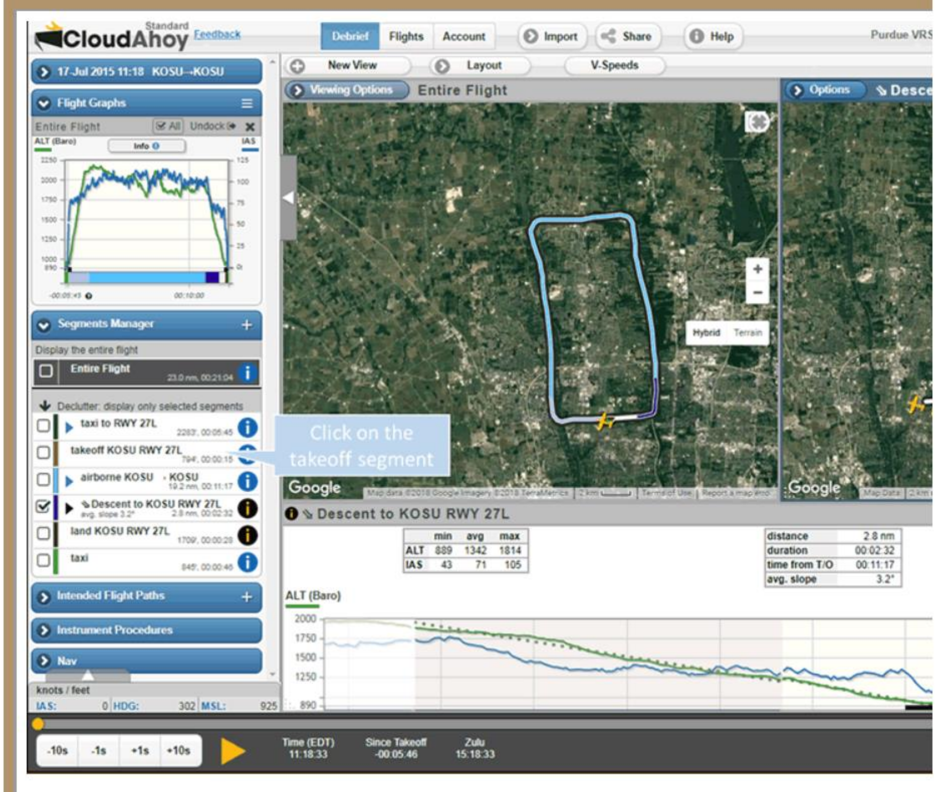
For each of the three flights, I created an interactive prototype debrief tool (Figure 19) using CloudAhoy screenshots and adding information on hazardous states. Respondents are able to interact with the screens to go back and forth between different displays, taking as much time as they need. The pilots have to pretend that this is a flight that they have just completed, and they have to answer the questions posed based on the information on the screens alone.

PURDUE UNIVERSITY

Review the following takeoff phase of flight as presented in these debrief screens, taking as much time as you need. The aircraft involved is a Cessna 172.

The debrief screens are semi-interactive: Under "Segments Manager," click on "takeoff KOSU RWY 27L" to choose the takeoff segment. Then click on each event you want to further investigate from the "takeoff safety information" list on the right.

When you are ready to answer questions about this takeoff, proceed to the next screen. Note that you will not be able to return to the debrief after clicking "Next."



Next →

Figure 19: Each flight has its own introduction and debrief screen, where pilots can interact with the debrief tool before continuing on to the questions.

5.2 Flights

I chose three flights to create debrief screens out of a larger set, based on the number of hazardous states in each flight. All flights originated at The Ohio State University Airport (KOSU) and were in a Cessna 172, an aircraft with which most pilots are familiar.

Flight A had four hazardous states—the aircraft rotated at a low airspeed while taking up too much of the runway, in a slight tailwind, and with a high centerline deviation. Flight B had a low Engine RPM and high airspeed at takeoff, but was otherwise safe. Flight C had a high centerline deviation and the takeoff took up too much of the runway. The airspeed at rotation was also slightly low. While the total risk of a flight cannot be measured with certainty, a simple risk metric is useful in comparing the three flights, as shown in
Table 6. I model the identified flight risk as a function of the hazardous states and triggers that were detected in the flight data, using a simple additive weighted model. This risk metric depends on not only the number of states and triggers, but also on their severity. For example, being 1 kt faster than recommended is not as dangerous as being 10 kts faster. To account for the various levels of risk, I divide each hazardous state into j qualitative degrees of risk, R_1, to R_j, in order of increasing severity. For example, deviating slightly from the runway centerline during takeoff but correcting for it quickly is a less hazardous state (R_1), but if throughout the takeoff the aircraft is increasingly deviating from the centerline, getting close to the edge of the runway, without any corrections, it becomes a more severe hazardous state (R_j).

I therefore model the total risk of a flight using Equation 1, where n is the total number of hazardous states and triggers that are being considered, j is the degree of risk for each hazardous state or trigger, k is the total number of risk levels chosen, and a_j is the weighting factor for each risk level.

$$Total\ Risk = \sum_{j=1}^{k} \sum_{i=1}^{n} a_j R_{ji} \quad (1)$$

For the takeoff example, I consider five hazardous states and three risk levels, as shown in Equation 2, where $a_1 = 1$, $a_2 = 3$, and $a_3 = 5$. These weighting factors increase the metric when higher risk hazards are present but still account for low risk hazards.

$$Total\ Takeoff\ Risk = \sum_{j=1}^{3}\sum_{i=1}^{5} a_j R_{ji} = \sum_{i=1}^{5} R_{1i} + 3\sum_{i=1}^{5} R_{2i} + 5\sum_{i=1}^{5} R_{3i} \quad (2)$$

I then scale the risk metric so that it can take values from 0 to 10, with 10 being the safer end of the spectrum, and 0 describing an unsafe flight. As shown in Equation 3, I divide by the risk metric of a hypothetical flight where all possible hazardous states and triggers occurred at an R_3 risk level.

$$Scaled\ Takeoff\ Safety\ Metric = 10 - 10 \times \frac{Total\ Takeoff\ Risk}{5\sum_{i=1}^{5} R_{3i}} \quad (3)$$

Table 6: I assigned scores of 1, 2, and 3 on the different hazardous states that were present in each flight based on how severe they were. Based on the final scaled safety metrics, Flight A is the riskiest one.

State	**Flight A**	**Flight B**	**Flight C**
Insufficient takeoff power	0	1	0
Inadequate/High airspeed at rotation	3	5	1
High tailwind or crosswind component	1	0	0
Insufficient runway distance remaining at takeoff	3	0	3
Deviation from centerline	3	0	3
Total Takeoff Risk:	**10**	**6**	**7**
Scaled Safety Metric:	**6**	**7.6**	**7.2**

5.3 Debrief Randomizer

Table 5 lists the eight possible ways of communicating risk messages in my full-factorial design experiment. For example, a hazardous state can be communicated in a graphical or numerical way. The two representations may bias the pilot differently, with one representation causing the pilot to think that a behavior is safer than another representation. Each factor in Table 5 has the potential to bias the pilot. Additionally, combinations of factors can affect pilots differently. For example, pilots may respond similarly to graphical and numerical methods, and risk-centric and safety-centric framing language methods, but at the same time respond differently to graphical methods that use risk-centric language. Table 7 shows a 2^3 full-factorial design with two-way interactions between the factors outlined in Table 5.

Table 7: The three factors result in eight possible combinations of factors that can be used in designing risk communication messages.

Combination	**Representation Method**	**Parameter Type**	**Framing Language**
1: [+1 +1 +1]	Graphical representation	Performance parameter	Safety-centric language
2: [+1 +1 −1]	Graphical representation	Performance parameter	Risk-centric language
3: [−1 +1 +1]	Numerical representation	Performance parameter	Safety-centric language
4: [−1 +1 −1]	Numerical representation	Performance parameter	Risk-centric language
5: [+1 −1 +1]	Graphical representation	Safety parameter	Safety-centric language
6: [+1 −1 −1]	Graphical representation	Safety parameter	Risk-centric language
7: [−1 −1 +1]	Numerical representation	Safety parameter	Safety-centric language
8: [−1 −1 −1]	Numerical representation	Safety parameter	Risk-centric language

The debrief randomizer function will randomly assign each pilot who takes the survey to one of the eight groups in Table 7. For example, in the first treatment combination, the debrief consists of graphical representations that use safety-centric language, and they describe parameters that refer to the system's performance, whereas in the last treatment combination, the debrief consists of are numerical representations that use risk-centric language and describe the safety of the system.

The first two factors in Table 7 (representation type and parameter type) are used in the debrief screens and messages, and the third factor (framing language) is used in the survey questions (Section 5.4).

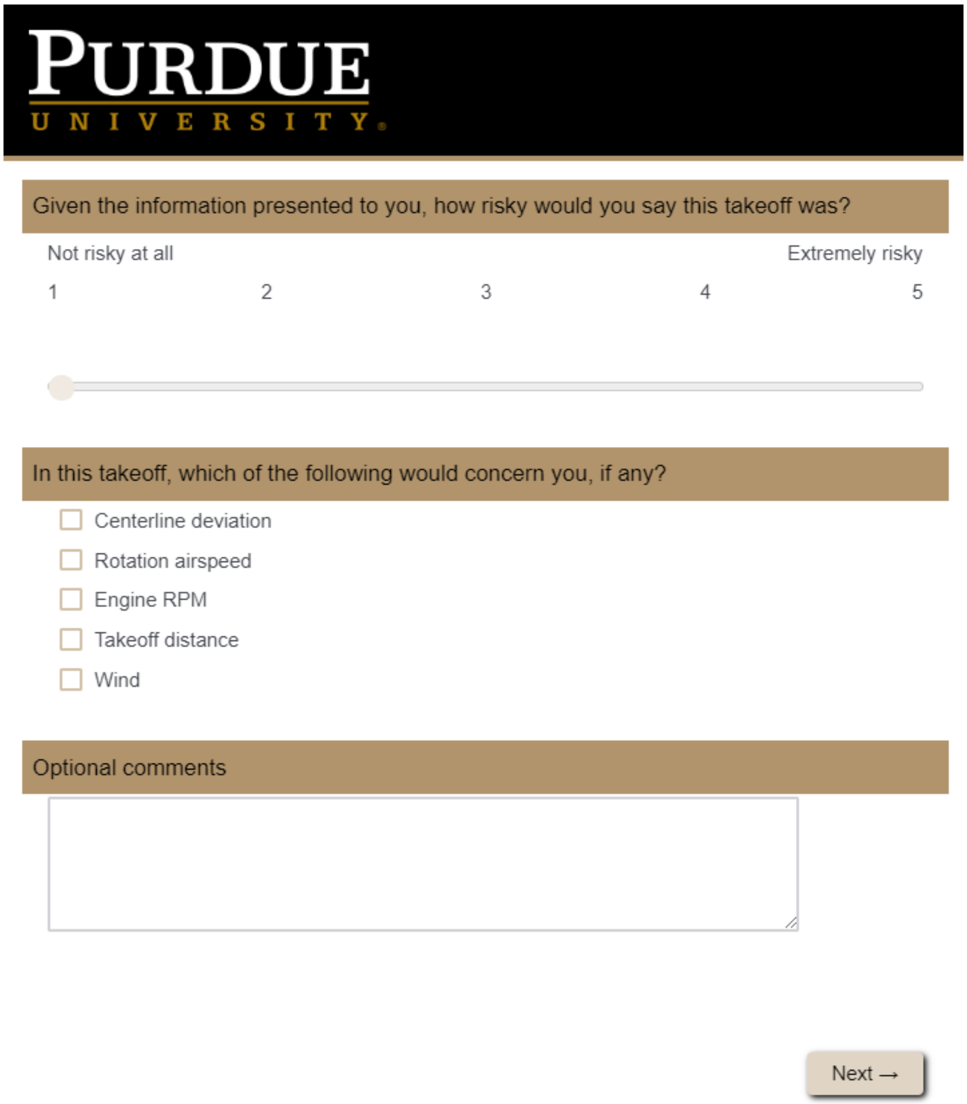
5.4 Survey Questions

As discussed in Chapter 3, I evaluate feedback effectiveness based on two characteristics: the accuracy of the perceived risk, and the pilot's willingness to change the identified unsafe behaviors.

5.4.1 Perceived Risk

The questions on the first post-debrief screen, shown in Figure 20, aim to address perceived risk. The first question, *Given the information presented to you, how risky would you say this takeoff was?* asks the pilot to rate the risk or safety of the flight on a 5-point Likert scale. Depending on whether the test is evaluating risk-centric language or safety-centric language, the question asks the pilots to use a risk scale or a safety scale, respectively. If different treatment combinations are affecting pilots' risk perception, then there will be a difference in the response distribution for Question 1 among different combinations.

The second question, *In this takeoff, which of the following would concern you, if any?* aims to investigate whether pilots identified the appropriate hazardous states in the takeoff. The pilots also have the opportunity to add comments.



PURDUE
UNIVERSITY

Given the information presented to you, how risky would you say this takeoff was?

Not risky at all — Extremely risky

1 2 3 4 5

In this takeoff, which of the following would concern you, if any?

- ☐ Centerline deviation
- ☐ Rotation airspeed
- ☐ Engine RPM
- ☐ Takeoff distance
- ☐ Wind

Optional comments

Next →

Figure 20: The first few questions address the risk perception part of the feedback effectiveness. The Likert scale changes depending on whether the respondent belongs to a risk-centric or safety-centric framing language treatment group.

5.4.2 Motivation to Change

To capture how likely pilots are to use the information provided to improve their future flights, I ask them to come up with changes that they could make to an upcoming flight, as shown in Figure 21. The three factors I am investigating could potentially impact the number of changes pilots recommend. Pilots may choose to say they would not make any changes.

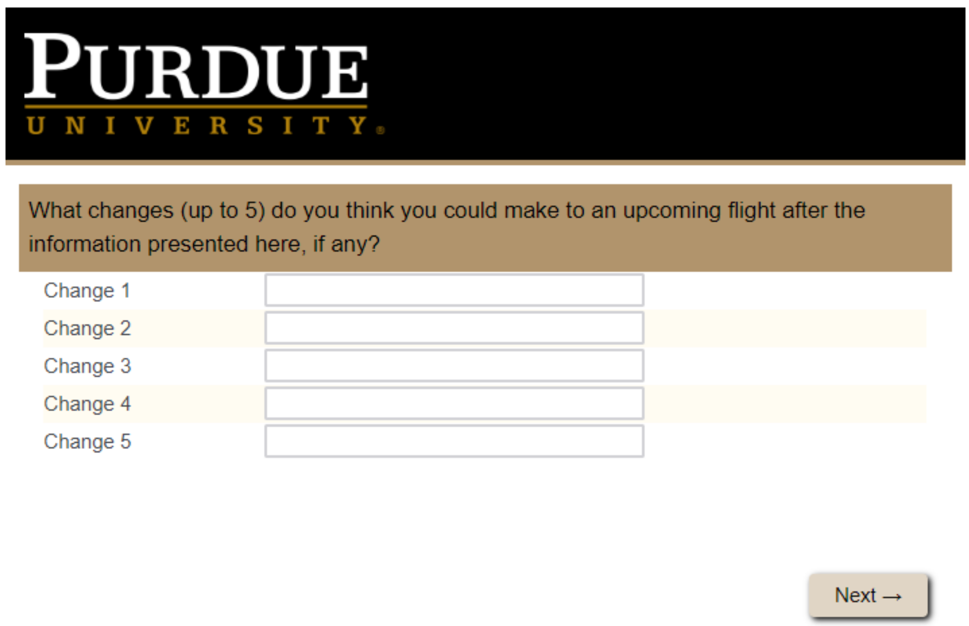


Figure 21: Respondents could suggest up to five changes that they could make to an upcoming flight after reviewing their debrief.

The answers to the question in Figure 21 feed into the next two questions: How likely are you to make each of these changes to an upcoming flight? and How important do you think each of these changes is to reducing risk on takeoff? shown in Figure 22. If the respondent does not indicate that they would make any changes, they are not presented with these questions. Effective risk messages, however, will motivate the pilots to change something that they have identified as being unsafe, so if the feedback is effective, I expect to see at least one of the changes rank highly on the 'likelihood' scale.

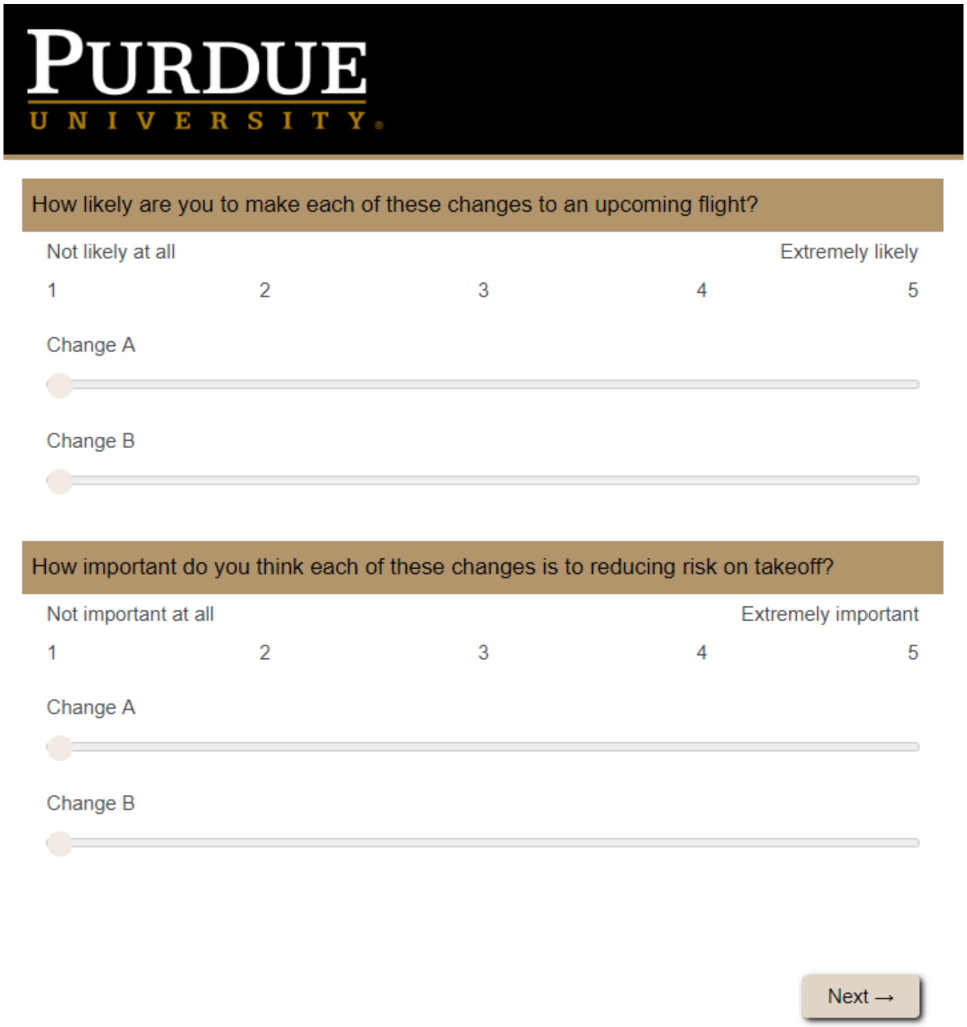


Figure 22: The respondents get two more Likert scale questions for each change they say they may make to an upcoming flight, asking about the likelihood of them actually making the changes as well as the importance of each change.

The respondents debrief three flights, and therefore answer these questions three times before moving on to the demographic questions.

5.5 Demographics

This section outlines the demographic questions that each respondent will have a chance to answer. None of them are mandatory, so some pilots may end up skipping all of them. However, since the demographic questions come at the end, the rest of the responses to the survey are still useful if a pilot decides not to participate in the last part. Most of these questions are multiple choice, with some of them requiring the respondent to type in a short answer, either numerical or a one-word response.

1. How old are you?

   18-24; 25-34; 35-44; 45-54; 55-64; 65 or older
2. What gender do you identify with?

   Male; Female; Other; Do not wish to specify
3. What is the highest level of education you have completed?

   Some high school; High school graduate or equivalent; Some college; 2-year degree; 4-year degree; Master's degree; Doctorate or Professional degree
4. What is your occupation?

   [Type in answer]
5. What kind of pilot's license do you currently hold?

   No certificate; Student; Sport; Recreational; Private; Commercial; Airline Transport
6. Which ratings or endorsements do you currently hold?

   Single-engine; Multi-engine; Instrument, Rotorcraft-Helicopter; Glider; Lighter-than-air; Seaplane; Complex; Tailwheel; High altitude; High performance; Flight instructor; Instrument flight instructor; Multi-engine flight instructor
7. How many years of flying experience do you have? Round to the nearest year.

   [Type in answer]
8. Was your flight training under Part 61 or Part 141?

   Part 61; Part 141; Both; I do not know
9. What kind of avionics do you most frequently use when you fly?

   Mostly steam gauges; Mostly glass cockpit; Both
10. How many flight hours do you have logged (approximately)?

    [Type in answer]
11. How often do you fly?

    2-7 days a week; Once a week; Once a month; Once every few months; Rarely/Never
12. How often do you participate in aviation safety programs and seminars (such as WINGS, FAASTeam seminars, etc.)?

    Monthly; 2-3 times a year; Once a year; Once every two years; Never
13. What is your home airport base? (ICAO identifier or city/state)

[Type in answer]

14. Have you used commercial debrief products, like CloudAhoy, before?

    Yes; No

6. SURVEY ANALYSIS

As described in Chapter 5, the experiment consisted of a full factorial design with three factors. The three factors are *representation method* (X_1), *parameter type* (X_2), and *framing language* (X_3). I want to ascertain the importance of each factor on Risk Perception (Y_1) and Behavior Change (Y_2). Each factor X_i has two levels, +1 and –1. For example, the Language factor, X_1, can take Risk-centric Language at the +1 level and Safety-centric Language at the –1 level. Since there are three factors, each run at two levels, there will be $2^3 = 8$ treatment combinations. Using a full factorial design allows me to investigate how multiple factors affect the output—risk perception and post-debrief behavior (Fala & Marais, 2019b).

The survey resulted in 187 responses that were entirely complete and a total of 268 usable responses. A usable response is any response where the pilot debriefed and responded to the questions of at least Flight A, whereas in a complete response the pilot has debriefed all three flights and then answered the demographic questions at the end. Since the survey consisted of three separate flight scenarios presented in the same order, I start by treating the three flights as distinct experiments. Flight A has 268 responses in total, Flight B has 195 responses, and Flight C has 189 responses. Table 8 shows the number of responses (replicates) for each treatment combination. Not randomizing the flights allowed me to maximize the number of responses for Flight A. improving the power of the tests on Flight A. Randomizing the flights would have improved the response rate on Flights B and C, with all flights resulting in approximately the same number of responses but would have also decreased the power for the analysis on Flight A. Not randomizing the three flights would provide equal representation for all flights in terms of responses, but getting an adequate power for the statistical analysis would require a high overall response rate.

The distribution of the overall risk perception data for Flight A (Figure 24, Appendix III) displays a bell-curve shape that is slightly skewed towards riskier values. The horizontal axis in Figure 24 consists of ordinal Likert-scale data from 1 to 5. The median and mean are close together, at 3 and 3.10 respectively, but while the ordinal scale used has a rank order, the intervals between values may turn out to not be equal (Jamieson, 2014). The overall distribution for the responses for Flight

B, shown in Figure 25 in Appendix III, is flatter and slightly skewed towards less risky values, with a standard deviation of 1.18. While the median (3) and mean (2.99) closely match those of Flight A, the responses tend to spread further away from the median. The distribution of Flight C responses (Figure 26, Appendix III) resembled the bell-curve of Flight A. Although Flight C has less responses, it maintained the central tendency around the neutral point, with the median (3) and mean (2.93) being very similar to those of Flights A and B. Overall, the pilots were able to understand the relative differences in risk between the three flights, but a lot of them chose to respond using the neutral option (i.e., 3 on the Likert scale), so the mode was 3 in all three flights. While I use non-parametric statistical analysis methods in this Chapter, I will also compare the results to methods used for normally-distributed continuous data because of the bell-curve shape of the data and the large sample size.

To determine how the three factors (representation method, parameter type, and framing language) independently impact the risk perception responses, I first created histograms of the response variable separated by factor levels. The difference in how the three factors influenced the results on the three flights suggests that the type of flight, level of risk, or type of hazardous states present in the flight could be additional factors that change how risk perception is affected.

In this Chapter, I first discuss the demographics and sample sizes that resulted from the survey, and then analyze the survey results using metrics that capture both facets of *risk effectiveness*. I use the answer to the question *"How risky do you think this takeoff was?"* to evaluate how two groups of pilots perceived their risk relative to each other, and the number of changes they report to the question *"What changes do you think you could make to an upcoming flight as a result of the information presented here, if any?"* The analysis is structured as follows: I first analyze the main effects of each of the three factors (representation method, parameter type, and framing language) on all metrics. I then analyze any interaction effects between the three factors and discuss the overall results.

6.1 Demographics

Out of all respondents, 188 worked on the demographics section. I deliberately designed the survey to transition through the three flight debriefs first before getting to demographics, so that I could still analyze data from respondents who decided to not finish the survey. Approximately 70% of the responses (i.e., those which had at least Flight A) were complete (had Flight A, B, C, and provided demographic information). Out of the respondents who provided demographic information, the majority were male (71% male and 26% female) and also completed at least a 4-year degree (76%). The majority of the sample consisted of private (49%) and commercial pilots (30%), with 58% of all respondents also having an instrument rating. Most respondents fly primarily aircraft with steam gauges (64%) and fly at least weekly (59%). The survey also reached pilots who are not as heavily involved in aviation—23% fly once a month, and 8% rarely fly. A few respondents (21%) reported that they never participate in aviation safety programs (such as the WINGS program, seminars, or training videos). Training for the largest portion of pilots was exclusively under Part 61 regulations (43%), while 13% of pilots trained exclusively through Part 141 schools, 31% in a combination of Part 61 and 141 programs, and another 13% did not know what kind of training they followed. Surprisingly, 88% had never used commercial debrief or flight visualization products like CloudAhoy before.

Table 8: As expected, not everyone completed the survey, which resulted in Flight A having more responses overall than Flights B and C. Most people stopped the survey after Flight A, with people being likely to complete the entire survey if they made it to Flight B.

Treatment Combination (X)	**Responses**		
	Flight A (268)	**Flight B (195)**	**Flight C (189)**
1: [+1 +1 +1]	35	29	19
2: [+1 +1 –1]	31	19	22
3: [–1 +1 +1]	44	24	28
4: [–1 +1 –1]	33	29	24
5: [+1 –1 +1]	23	22	23
6: [+1 –1 –1]	34	21	19
7: [–1 –1 +1]	33	26	27
8: [–1 –1 –1]	35	25	27

6.2 Representation Method

The representation method factor could take one of two levels: graphical, or numerical. Table 9 shows the number of responses corresponding to each level of the representation method factor for each flight. Although the survey software presented the graphical and numerical representation type versions of the survey to equal numbers of respondents, noticeably fewer respondents completed the graphical version of the survey, for all three flights. This discrepancy potentially suggests that pilots prefer to review numerical data, instead of trying to decipher graphical information. A chi-square test does not support the hypothesis that the completion ratios among graphical representations and numerical representations are different. The median pilot took 89 seconds to review their debrief in Flight C for both representation types, 116 seconds and 95 seconds in Flight B for graphical and numerical representation methods respectively, and 49 seconds and 60 seconds in Flight A for graphical and numerical representation methods respectively.

Table 9: The responses were split unevenly between the graphical and numerical levels. Overall, the number of responses decreased with each flight, and more people responded to the numerical representation version of the survey. The percentages represent the completion ratio among pilots who saw the specific survey version. For example, 368 pilots were presented with the graphical representation method version of the survey for Flight A, and 123 completed it, resulting in a 33% completion percentage. The completion ratio increased with each flight.

	Number of responses					
Flight	**Graphical**		**Numerical**		**Total**	
A	123	33%	145	39%	268	36%
B	91	64%	104	73%	195	68%
C	83	77%	106	98%	189	88%

Figure 30 in Appendix III separates the responses of pilots who debriefed the flight graphically and numerically for all three flights in six histograms. The representation method factor changed the response mode only in Flight B. Flight C appeared to be largely unaffected. Flight A maintained the same mode, but the graphical representation was more uniform in distribution around the mid-point than the numerical level. Flights A and B seem to have moved in opposite directions—the graphical representation moved the responses slightly towards the riskier side in Flight A compared to the numerical representation, but distinctively towards the less risky side in Flight B. Table 10 shows some of the descriptive statistics on risk perception for the graphical and numerical representation methods among the three flights. Section 5.2 discussed one way to calculate the risk in a takeoff based on the number and severity of the hazardous states present in it. Using that metric, Flight A is the riskiest takeoff, and Flight B the safest takeoff. This ranking appears in the mean risk perception score for the graphical representation, but the highest mean in risk perception across all three flights for the numerical representation suggests that the pilot sample thought that Flight B was the riskiest takeoff among the three flights.

Table 10: The way pilots rated their risk changed between graphical and numerical representation methods. The largest change happened in Flight B, with the mean increasing from 2.76 in the graphical method to 3.20 in the numerical method.

	Risk rating							
	Graphical				**Numerical**			
Flight	**Mean**	**Standard Deviation**	**Median**	**IQR**	**Mean**	**Standard Deviation**	**Median**	**IQR**
A	3.1951	0.9889	3	2	3.0138	1.0340	3	2
B	2.7582	1.2679	3	2	3.2019	1.0647	3	2
C	2.9277	0.9342	3	2	2.9340	1.0353	3	2

The number of changes the pilots suggested after reviewing their feedback is an indication of how motivated they are to change unsafe behaviors. Overall, the safer takeoff in Flight B resulted in a lower number of suggested changes, and the riskier takeoff in Flight A in more changes, as shown in Table 11. The riskier takeoff also resulted in pilots suggesting more changes they would make to an upcoming flight. Figure 31 in Appendix III shows the effect of representation method on the number of changes the pilots came up with in each of the three flights. Although the numerical representation did decrease the number of respondents who opted for no changes after their debrief in Flights B and C, the same did not apply to Flight A, the riskier flight. In Flight B, in particular, the numerical representation resulted in most respondents saying they would make two changes, whereas the mode for the graphical representation method was zero. The average response among pilots was a total of 50 (graphical representation) and 52 (numerical representation) characters in length for Flight A, as shown in Table 33 in Appendix III, 36 and 53 for Flight B, and 51 and 45 for Flight C.

Table 11: The number of changes pilots suggested after reviewing their debrief ranged from zero to five. The numerical representation method resulted in a higher number of changes overall.

	Number of changes							
	Graphical				**Numerical**			
Flight	**Mean**	**Standard Deviation**	**Median**	**IQR**	**Mean**	**Standard Deviation**	**Median**	**IQR**
A	1.3984	1.3474	1	2	1.5724	1.3629	2	3
B	1	0.9661	1	2	1.3654	0.8251	1	1
C	1.3133	1.1575	1	2	1.5377	1.0882	2	1

The Mann Whitney U test showed statistical differences in the numerical and graphical distributions for the risk perception in Flight B but not Flights A or C. I used the Mann Whitney

U Test because the survey data consists of a categorical independent variable (representation method) of two levels (numerical and graphical) and an ordinal dependent variable (Likert-scale data).

The results suggest that choosing a numerical representation when communicating risk to pilots can potentially push them towards seeing a higher level of risk in their flights and also make them more willing to change their behaviors. Pilots also had a greater tendency to give up on the graphical methods, perhaps being frustrated at having to decipher information. These findings suggest that the risk communication biases in other fields may not be applicable to aviation—in medicine, patients prefer graphical methods, often due to a lack of numerical literacy, which is not the case among pilots.

6.3 Parameter Type

The parameter type factor can also take one of two values: safety parameters, or performance parameters. Parameter type refers to whether the parameter is presented in terms of risk or performance. For example, comparing the amount of runway that the pilot used in taking off to the takeoff distance specified in the aircraft handbook, tells the pilot how close they were to the nominal way of flying in comparison to the handbook. Reporting the amount of runway that remained after takeoff instead (i.e., the runway length that was not used) aims to communicate how much room for error the pilot had, based on how close the pilot and aircraft are to an unsafe situation or incident. In the first case, the pilot should want to minimize the number; in the latter case, a higher number is better. Table 12 shows the number of responses corresponding to each level of the parameter type factor for each flight. In this parameter, there is no consistent discrepancy between number of responses for the two levels.

The median pilot took 88-99 seconds to review their debrief in Flight C for both parameter types, 106 seconds and 102 seconds in Flight B for performance parameters and safety parameters respectively, and 60 seconds and 46 seconds in Flight A for performance and safety parameters respectively. Similar to the results for representation method, the respondents in Flight C took more time to review their debrief and the time responses are less affected by outliers and the people who decided not to continue.

As shown in Table 12, the completion rates were more even among safety and performance parameters. The difference in completion rates between safety and performance parameters was less than or equal to 5% in all three flights.

Table 12: The responses were split more evenly between the performance parameter and safety parameter levels compared to the representation method factor.

	Number of responses					
Flight	**Performance**		**Safety**		**Total**	
A	143	39%	125	34%	268	36%
B	101	71%	94	66%	195	68%
C	93	86%	96	89%	189	88%

Figure 32 in Appendix III separates the risk perception responses of pilots who debriefed the takeoffs in terms of safety parameters and performance parameters for all three flights. The distribution of risk perception responses in Flight A had a slightly smaller variance in the safety representation. The safety parameters in Flight B moved the responses to the right, towards *extremely risky*, with a different mode in the *safety* and *performance parameters* cases. The parameter type made no noticeable difference in Flight C. Both Flight A and C maintained the same mode.

Table 13 shows some of the descriptive statistics on risk perception for the performance and safety parameter types among the three flights. According to the risk metric, Flight A is the riskiest takeoff, and Flight B the safest takeoff. The means among the three flights for the *performance parameter* correspond to the risk metric ranking, but the mean for the *safety parameter* for Flight B suggests that it would be the riskiest takeoff among the three flights.

Table 13: Pilots risk perception changed between performance and safety parameter types in Flight B, with the mean increasing from 2.63 in the performance parameter type to 3.38 in the safety parameter type.

	Risk rating							
	Performance				**Safety**			
Flight	**Mean**	**Standard Deviation**	**Median**	**IQR**	**Mean**	**Standard Deviation**	**Median**	**IQR**
A	3.1538	1.0299	3	2	3.0320	0.9995	3	2
B	2.6337	1.0145	3	1	3.3830	1.1742	3	1
C	2.9570	0.9659	3	2	2.9063	1.0165	3	2

Figure 33 (Appendix III) shows the distributions of the number of changes pilots suggested after reviewing their debrief for all three flights, split by parameter type. The *safety parameter* increased the number of changes pilots suggested after reviewing their feedback in Flights B and C, but decreased the mean of the number of changes in Flight A, as shown in Table 11. For the riskier takeoff in Flight A, pilots suggested more changes they would make to an upcoming flight when presented with the *performance parameters*. The safety parameter version of the debrief reduced the respondents who opted to continue without making any changes in Flight B and C (the safer takeoffs), but increased the "no changes" responses in Flight A (the riskier takeoff). The average response among pilots was a total of 56 and 46 characters in length for performance and safety parameters respectively in Flight A, as shown in Table 33 in Appendix III, 37 and 54 in Flight B, and 48 and 59 in Flight C. The discrepancy between the two groups is higher in the *parameter type* factor than the *representation method* factor.

Table 14: The numerical representation method resulted in a higher number of changes overall.

	Number of changes							
	Performance				**Safety**			
Flight	**Mean**	**Standard Deviation**	**Median**	**IQR**	**Mean**	**Standard Deviation**	**Median**	**IQR**
A	1.6364	1.3612	2	3	1.3280	1.3367	1	2
B	1.000	0.8602	1	2	1.4043	0.9196	1	1
C	1.3226	1.0951	1	2	1.5521	1.1413	1.5	1

Similarly to the *representation method*, the Mann Whitney U test showed statistical differences in the performance and safety parameter type distributions for the risk perception in Flight B (p-value 6.9202e-06) but not Flights A or C (p-value 0.4316 and 0.7806 respectively).

From the results presented in this section, it is not clear whether risk communication for pilots should use performance parameters, safety parameters, or a combination of the two types. *Performance parameters* encouraged pilots in Flight A to rank the takeoff as being riskier and suggest more changes, whereas *safety parameters* did the equivalent for Flight B.

6.4 Framing Language

Figure 34 in Appendix III shows how pilots responded when asked "*How risky would you say this takeoff was?"* versus "*How safe would you say this takeoff was?*" While mathematically a flight that is not too risky is by definition very safe, the phrasing did affect how pilots perceived the risk. Although Flight C was not affected by the framing language, as with the other two factors, in the safety-centric language there is a higher concentration towards the neutral value. However, the framing language factor changed the mode in Flight B and altered the distribution in Flight A. Flight B also shows a general movement towards the riskier side when using a safety-centric framing language.

Table 15 shows the number of responses corresponding to each level of the parameter type factor for each flight. In this parameter, the difference between the number of responses for the two levels is small and inconsistent among the three flights. The difference in completion rates between safety and performance parameters was less than or equal to 5% in all cases.

Table 15: The complete responses were split evenly between the *safety-centric language* and *risk-centric language* levels compared to the *representation method* factor.

	Number of responses					
Flight	**Safety-centric**		**Risk-centric**		**Total**	
A	135	37%	133	36%	268	36%
B	101	71%	94	66%	195	68%
C	97	90%	92	85%	189	88%

Figure 34 in Appendix III separates the risk perception responses of pilots who debriefed the takeoffs in a *safety-centric* language and in a *risk-centric* language for all three flights. The framing language did not make a difference in Flight C. The impact on Flight A was different than the impact on Flight B—in the first case, framing the question in safety-centric language resulted in more people saying the flight was not safe, whereas in the second case framing the question in risk-centric language decreased the number of people saying the flight was not risky and increased the number of people reporting that the flight was risky.

Table 13 shows some of the descriptive statistics on risk perception for the risk-centric and safety-centric framing languages among the three flights. The framing language resulted in a change in the mean in Flight A but did not affect Flights B or C. While *framing language* changed the mode in Flight B, it did not change the mean, so an ANOVA could not report that change since it only compares the means.

Table 16: The framing language caused a slight change in the means of the risk perception in Flight A.

	Risk rating							
	Safety-centric				**Risk-centric**			
Flight	**Mean**	**Standard Deviation**	**Median**	**IQR**	**Mean**	**Standard Deviation**	**Median**	**IQR**
A	3.2519	1.0631	3	1.75	2.9398	0.9436	3	2
B	2.9802	1.0953	3	2	3.0106	1.2742	3	2
C	2.9897	0.9947	3	2	2.8696	0.9858	3	2

Figure 35 (Appendix III) shows the distributions of the number of changes pilots suggested after reviewing their debrief for all three flights, split by framing language. The distributions look the same for all flights independent of the framing language used. Table 14 shows the descriptive statistics for all three flights. The average response among pilots was a total of 55 and 50 characters in length for the safety-centric and risk-centric framing languages respectively in Flight A, as shown in Table 33 in Appendix III, 46 and 44 in Flight B, and 51 and 55 in Flight C. The difference between the two framing languages is much lower in this factor than the *representation method* and *parameter type* factors.

Table 17: The framing language did not change the number of changes that the pilots said they would make to an upcoming flight.

	Number of changes							
	Safety-centric				**Risk-centric**			
Flight	**Mean**	**Standard Deviation**	**Median**	**IQR**	**Mean**	**Standard Deviation**	**Median**	**IQR**
A	1.5556	1.3308	2	2	1.4286	1.3833	1	3
B	1.1486	0.9098	1	2	1.2447	0.9121	1	1
C	1.4536	1.1816	2	2	1.4239	1.0611	1	1

The Mann Whitney U test showed statistical differences in the safety-centric and risk-centric framing languages for the risk perception in Flight A (p-value 0.0093) but not Flights B or C (p-value 0.9770 and 0.4758 respectively).

From the results presented in this section, I cannot conclude that debrief should be using one particular framing language over the other, since framing language was only significant in one of the flights (the riskiest takeoff), and additional experiments are required.

6.5 Interaction Effects

I ran the full complement of all possible factor combinations to estimate all of the main effects between the factors and the results, as well as any interaction effects between factors. The full factorial design has three main effects and three two-factor interactions.

To test for interaction effects, I used the Scheirer-Ray-Hare non-parametric test for each flight. Because the Scheirer-Ray-Hare test is used for a two-way factorial design, I ran it with three combinations of two-factor pairs (Table 21, Appendix III). For Flight A, this test identified the *framing language* factor as a main effect on the risk perception response and the *representation method* and *parameter type* factors together as an interaction effect. The *parameter type* factor was a main factor that impacted the number of changes the pilots recommended in Flight A. A three-way ANOVA identified the same main and interaction effects (Table 27, Appendix III). Neither the Scheirer-Ray-Hare test (Table 22, Appendix III) nor the ANOVA (Table 28, Appendix III) identified any main effects or interaction effects which impact the number of changes pilots suggest after reviewing their debrief. In Flight B, both the Scheirer-Ray-Hare test (Table 23 and Table 24 in Appendix III) and the ANOVA (Table 29 and Table 30 in Appendix III) identified *representation type* and *parameter type* as main factors in both risk representation and the number of changes pilots suggested they would make. There were no significant main or interaction effects in Flight C (Table 23, Table 24, Table 31, and Table 32 in Appendix III).

Since the flight appears to be a factor in how different representation methods affect pilots, I ran a four-way ANOVA (Table 18) for *representation type* (two levels), *parameter type* (two levels), *framing language* (two levels), and *flight number* (three levels) as factors, considering all data points together rather than separating by flight. Parameter type affects responses significantly across all three flights. The performance parameter type results in pilots rating their risk as lower. Framing language also impacts pilots across all three flights, although not at the 5% significance level. Safety-centric language results in pilots rating their risk as higher. There is an interaction effect between representation type and parameter type across the board. The flight interacts with the representation method and parameter type factors, suggesting that representation type and parameter type may impact pilots differently depending on the flight.

Table 18: *The ANOVA test indicates that parameter type (and framing language, to a lesser extent) affect how pilots rate their risk across all three flights. The parameter type interacts with the representation method. The flight interacts with representation method and parameter type but not with framing language.*

Source	**Sum Sq.**	**d.f.**	**Mean Sq.**	**F**	**Prob>F**
Repres	0.829	1	0.8287	0.78	0.3778
Param	7.329	1	7.3286	6.89	0.0089
Lang	3.284	1	3.2837	3.09	0.0794
Flight	3.238	2	1.6188	1.52	0.2192
Repres*Param	4.903	1	4.9033	4.61	0.0322
Repres*Lang	0.762	1	0.7615	0.72	0.3979
Repres*Flight	11.35	3	5.6748	5.33	0.005
Param*Lang	0.086	1	0.0861	0.08	0.7762
Param*Flight	22.024	2	11.0121	10.35	0
Lang*Flight	2.583	2	1.2915	1.21	0.2978
Error	677.797	637	1.064		
Total	33.779	651			

6.6 Proportional Odds Model

The proportional odds model is a regression model for ordinal dependent variables, such as the risk categories in the survey, where “extremely risky” is riskier than “somewhat risky” but the difference between the two categories cannot be numerically quantified. Since respondents could choose one of five categories to characterize the risk of the flight, there are four logarithms of the odds of answering in certain ways:

$$\ln\left(\frac{P(Y \leq i+1)}{P(Y > i+1)}\right) = a_i + \sum_{j=1}^{n} \beta_j X_j \tag{4}$$

where $i = 1, 2, 3, 4,$ corresponds to the four logarithms, and $j = 1, 2, 3, 4,$ corresponds to the four factors as described in the ANOVA in Table 18. The model in Equation 4 has four different intercepts but common coefficients (slopes) among the different categories. Table 19 lists the four coefficients for the *logit* function and their corresponding p-values. The proportional odds model identified the same factors as the parametric tests that assumed continuous variables.

Table 19: The proportional odds model shows that the parameter type and framing language factors are significant at the 5% significance level. The flight does impact the pilot's risk perception, but not at the same level as the parameter type and framing language factors.

	β_i	p-value
β_1 (Representation method)	0.0972	0.4930
β_2 (Parameter type)	0.3115	0.0281
β_3 (Framing language)	-0.2986	0.0350
β_4 (Flight)	0.1428	0.0944

6.7 Qualitative Responses

The word cloud in Figure 23 corresponds to the changes pilots said they would make to an upcoming flight in Flight A. The words that appeared most frequently in the responses were *centerline* (83 times), *runway* (79 times), and *wind* (69 times). The words *proper* and *speed* usually went together. *Takeoff* and *rotate/rotation speed* also ranked high. *Tailwind*, *crosswind*, and *RPM* appeared lower on the list. The changes the pilots suggested mapped to the hazardous states that were present in the flight. The word clouds for Flight B and C are shown in Figure 36 and Figure 37 respectively (Appendix III).

The average likelihood value and maximum likelihood value are the average response and maximum response of each pilot out of all their changes to the question *How likely are you to make these changes to an upcoming flight?* respectively. The maximum likelihood value should ideally be high if the feedback was effective, because it has motivated the pilot to do something to improve their flying on their next flight. The average likelihood value is dependent upon the number of changes that the pilot has provided—if they provide one change that they rate high and four changes that they rate low, the average likelihood value will be lower than a pilot's who provided one highly-rated change. Both the average and maximum value distributions were similar across the three flights and all responses ranked highly. Each pilot suggested changes that they were likely to make (4 or 5 on the Likert scale).

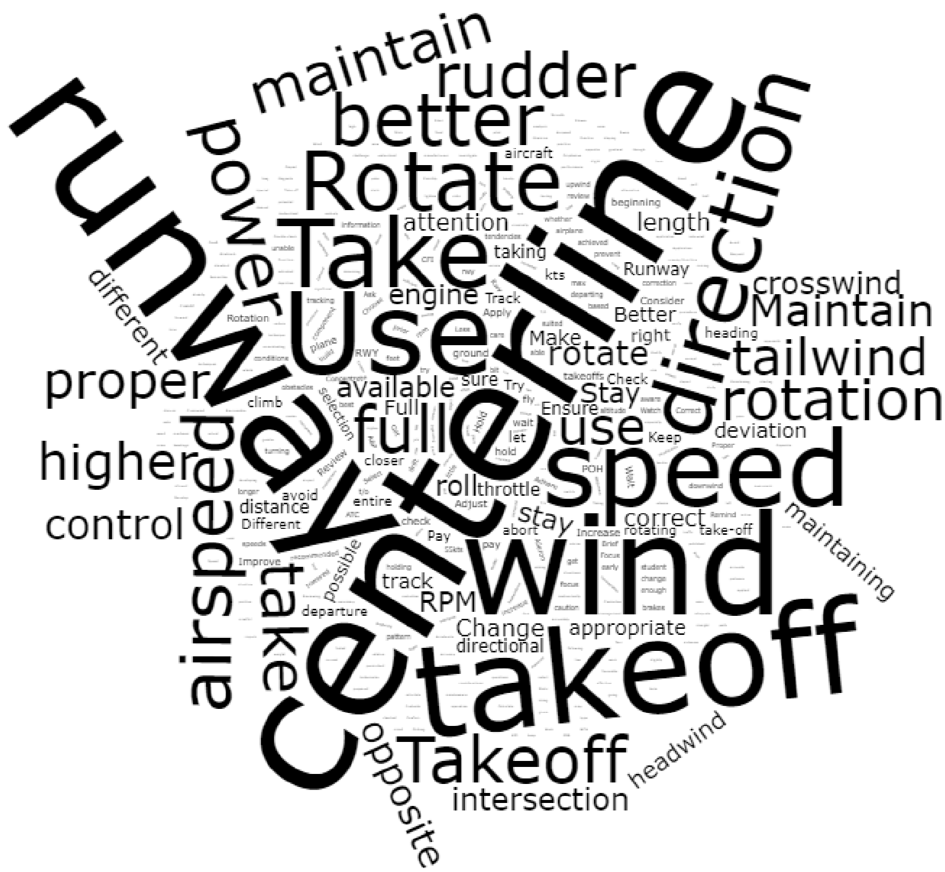


Figure 23: The respondents to Flight A reported that they would make a total of 411 changes to an upcoming flight after reviewing their debrief.

6.8 Survey Analysis Results and Discussions

The analysis of the survey has shown that debrief representation affects risk perception, but not necessarily willingness to change unsafe behaviors. Overall, respondents appeared to be highly motivated to change the behaviors they identified as unsafe, independent of how the information was presented to them. Pilots reporting that they are likely to make the changes they reported, however, could potentially be a result of self-selection bias, where the pilots who took the survey already care about improving their safety. Using scenarios instead of data from flights that the pilots have actually flown might have had an impact on how willing the pilots were to recommend changes, suggesting that the flight was not entirely safe. All three factors investigated had an effect on risk perception, although not for the same flight. Framing language affected risk perception in Flight A, whereas parameter type and representation method affected risk perception in Flight B. Flight C remained unaffected by all three factors. The discrepancy in how the three factors impact risk perception in different flights may be a function of the risk level of each flight (Flight A is riskier than Flights B and C, as discussed in Section 5.2) or the kinds of hazardous states present in each flight. An additional survey would be required to investigate the effect of the flight characteristics on risk perception.

These results suggest how flight instructors could change how they debrief a student's flight, based on what the instructors want to achieve. Overall, the graphical representation method caused pilots to report a lower risk for the same flight, compared to the numerical representation method. If a flight instructor wants to encourage a student to practice more, they could take advantage of the representation method bias—using numerical representation formats would likely bias the student towards perceiving the flight to be riskier and wanting to make more changes to an upcoming flight. Similarly, using safety parameters instead of performance parameters can help students think more about their risk and how to improve. For example, instead of focusing on how close the airspeed is to the appropriate airspeed or manufacturer's recommendations, the instructor can focus on how close the airspeed is to an unsafe situation.

Based on the interaction effects in Table 18, there is at least one more factor that affects risk perception. The flight itself affected how pilots were biased by the three factors. The concealed factor could be the risk present in the flight (meaning that we should be using different representation methods for safer flights than riskier flights) or the specific events in question (e.g., perception of airspeed deviations and centerline deviations may be biased differently by the three factors).

Lastly, parametric tests such as ANOVA or tests meant for continuous or categorical variables may be unreliable when used with ordinal data. However, Likert-scale data displays traits of continuous variables and researchers may treat them as such if they consist of five or more categories without harm to the analysis, referring to the variable as an ordinal approximation of a continuous variable (Sullivan & Artino, 2013). For the purposes of this experiment, all tests performed provided similar results in terms of which factors significantly bias pilot risk perception and motivation to change, supporting the ordinal approximation of a continuous variable argument.

7. CONCLUSION AND FUTURE WORK

Safety has been at the forefront of the aviation industry ever since its conception. The continuous attempts at improving the safety record of flight have made air travel a relatively safe transportation mode. When accidents do occur, the lessons learnt are implemented in the rest of the fleet as well as future designs. The first fatal powered-aircraft crash in 1908 resulted in the first Army pilots wearing helmets, which ensured they would not die in the same way as the one historical data point available. However, GA safety still lags, and the expected growth in the aviation industry makes improving the safety record imperative.

There are numerous challenges in GA that make reactive safety measures less effective than in commercial aviation—the large diversity in the aircraft fleet and equipment, pilot experience, training, and qualifications makes it difficult to employ generalized methods across the board. In this research, I focused on improving safety by reducing human error through safety-driven debrief. Effective safety-driven debrief has to communicate the risk to pilots and motivate them to mitigate any unsafe behaviors. A comprehensive treatment to the topic should consider the question *What is the best way to communicate unsafe behaviors effectively to pilots?* To achieve this goal, we first need to answer the following two questions *1) Which unsafe behaviors should be communicated to pilots?* and *2) How do we measure and calculate unsafe behaviors in different forms of flight data?*

In this research, I addressed the three aforementioned questions. First, I **identified unsafe events** that need to be communicated to pilots by using the NTSB database to create a list of hazardous states and triggers. For the purposes of this thesis, I **created a list of hazardous states that appear in the takeoff phase of flight**, which allowed me to have a more bounded list of hazardous states. Communicating hazardous states that have appeared in accidents either as causes of the accident or as factors ensures that the behaviors I am asking the pilots to correct are in fact unsafe. In this research, I identified the factors that appear in takeoff accidents and also mapped them to private pilot certification standards (ACS) to create the list of events/unsafe behaviors that should be communicated.

I then described **measurable parameters that can be reasonably mapped to the hazardous states** during the takeoff phase and **developed algorithms** to detect them in flight data. This research focused on FDR data, although the same techniques can be applied to other sources of data, with higher uncertainty in the final outcome. In particular, I calculated wind components at takeoff, runway distance remaining (and runway distance used), and deviation from the centerline (and distance from the edge of the runway) by combining different sources of data as described in the tables of Chapter 3.

The overall research question aims to investigate how to best communicate feedback to pilots, based on their flight data. I defined feedback effectiveness based on whether it communicated the risk of the situation and whether it motivated pilots to improve their flying by doing something to mitigate the unsafe behaviors. To answer this question, I **created graphical and textual/numerical representations** to communicate each hazardous state, both in terms of a safety parameter and a performance parameter when applicable. I put these representations in a debrief format based on CloudAhoy and created an interactive prototype for a debrief tool to be used in a survey. I **created and disseminated a survey** based on a full-factorial design to evaluate whether three factors I chose based on the literature on cognitive biases in risk communication in other fields affect feedback effectiveness in aviation. I analyzed the results from 268 responses and showed that the feedback representation does affect its effectiveness in terms of risk perception, but not when it comes to pilots' motivation to change. The effect of the three factors is not consistent across the three flights. Section 7.1 discusses the limitations of the work and the results, and Section 7.2 suggests future work that can mitigate the limitations and advance the research.

As discussed in Chapter 4, the risk communication community has studied how to decrease risk among the general public and in our daily lives by helping people take risk seriously—how to convince people to use their seat belts in a moving vehicle, quit smoking, not eat raw eggs. In all these situations, there are specific actions that the public can take to mitigate the risks, and the actions could even be legislated. We therefore have some idea of how to address and prevent violations, or decision errors. However, helping people manage risk when the actions don't consist of simple rules, i.e. when the errors are skill-based, is proving to be more difficult, and the literature does not address the risks people take in the workplace through their own performance. In the aviation industry, we can apply the research from the literature to help pilots wear their seatbelts, use their checklists, and check their fuel before they take off. However, we do not know how the way we talk to pilots about their performance and risk affects them. This research **identified how pilots can be biased by different communication and risk representation methods** and **showed how to uncover information** with regards to these biases.

7.1 Limitations

There are certain limitations to this work that arise mainly as a result of the survey design. Firstly, the post-debrief questions are not mandatory. There is no way of knowing whether respondents who chose to not recommend any changes to make are making that choice deliberately or out of time constraints. The results were not the same for all flights, suggesting that the flight may be one of the factors that affects feedback effectiveness. To eliminate the effect of the flight type on feedback, it would perhaps be beneficial to present pilots with feedback on specific states, instead of the flight as a whole. While narrowing the survey down to particular states would take away the realism of a flight debrief, it would also decrease the number of variables. Other researchers may also disagree with the communication factors that I investigated, arguing that other factors may also influence feedback effectiveness, but studying more factors simultaneously results in a lengthier survey which requires more responses from which to draw conclusions.

Another limitation was the type of survey data collected—Likert-scale data is ordinal and makes statistical analysis difficult. Pilots also interpret Likert-scale data independently, and it is not possible for researchers to know how the pilots interpreted them. Additionally, even though I followed survey design guidelines from the literature in creating the survey for this work, we do not know how pilots respond to different kinds of survey questions. For example, pilots were able to choose the *neutral midpoint* in each Likert-scale in the survey. Not having the midpoint as an option would have pushed pilots to choose the *safe* or *risky* side of the scale.

In general, it is inherently difficult to draw conclusions from the pool of pilots who took the survey, because we do not know how they answered the questions. It is possible that respondents misunderstood questions or did not truthfully answer, since the questions were not based on their own flights. Using focus groups or in-person simulations and questioning would remedy the uncertainty that comes with an anonymous survey, but it would decrease the sample size and increase the pilot workload.

7.2 Recommendations for Future Work

In this research, I took the first steps towards understanding the cognitive biases that come into play when communicating risk-related information to pilots. This section discusses how each of the three questions can be further expanded to increase the impact of the work.

7.2.1 Hazardous States During Other Phases of Flight

The same techniques can be applied to hazardous states that occur during phases of flight other than takeoff. For example, a different set of states is important during a go-around, such as pitch attitude, angle of attack, airspeed, aircraft configuration, etc. Applying the work on a different phase of flight can show whether the same observations and discussion in Chapter 6 apply to all states or if the observations are state-specific.

7.2.2 Application of State Detection Algorithms to Data of Lower Resolution

In this thesis, I focused on FDR data, mainly provided from a Garmin G1000 flight deck system. As discussed in Chapter 2, it is possible to repeat this work with data of lower resolution, such as smartphone data. Data of lower resolution or quality would introduce increased uncertainty in all parts of Figure 7 in Chapter 2. Firstly, the lack of some parameters, such as engine RPM or airspeed, makes the identification of phases of flight more complicated and less accurate. For example, in the takeoff case, I would have to identify the takeoff roll based on altitude (recorded with less accurate sensors than the aircraft instrumentation) and forward velocity alone, instead of using altitude in combination with engine RPM, pitch attitude, and vertical speed. Then, I would also have to detect hazardous states using parameters that I calculate with less accuracy. For example, in this thesis I was able to calculate deviation from the runway centerline precisely using G1000 data. To detect the same hazardous state using smartphone data, I would first have to find an approximate takeoff point, and then calculate an approximation to the centerline deviation, which could push the flight into a false alarm hazardous state prematurely.

The higher uncertainty in the detection of hazardous states could potentially make the pilots less likely to believe the debrief and less likely to change their behaviors, so the results of Chapter 6 may not necessarily still apply with data of lower resolution.

7.2.3 Investigating How Pilots Respond to Surveys

In designing the survey for this research, I used literature on survey-biases that is generalized on the population as a whole. Guidelines included recommendations for when to use the neutral point in Likert scales and the implications of anchoring, among other biases. However, no one has investigated whether pilots respond to surveys in the same way as the general population or if they are subject to the same kinds of biases.

A different kind of survey, designed to test different survey-taking biases with a control group and a number of experimental groups, can shed light on whether pilots are affected by different survey design techniques. Some of the factors that this survey can experiment with are scale length (for example, a 5-point Likert scale versus a 10-point Likert scale), continuous versus discrete scales, and the effect of excluding the neutral point (for example, not including the middle value 3 on a 5-point Likert scale) on rating scales.

The results of such a survey could help inform and increase the reliability and impact of future aviation research that implements other surveys.

7.2.4 Effect of Demographics

In this research, I investigated the effect of three communication factors on risk perception and motivation to change among all respondents. However, I collected a plethora of demographic information through the survey that can be used to investigate whether there are pilot characteristics that influence how pilots perceive their risk or understand their debrief. Some of these characteristics could be gender, age, flight experience, flight training background (Part 61 versus Part 141), education level, or occupation.

Using the demographic information collected limits the number of responses that could be considered, as a number of pilots stopped taking the survey before reaching the demographic questions.

7.2.5 Using Simulator Studies in Place of the Survey

In the survey, I asked pilots to treat the flights they were debriefing as scenarios and pretend they were the ones flying, as it is easy to be more critical of a flight when we are not the ones responsible for it. Automating the creation of debrief screens and the tool could enable a more real-time application of the research, with participants completing a flight on a simulator before looking at their own data, answering questions on risk perception and willingness to change behaviors in a more realistic way.

APPENDIX A. FDR DATA LONG COMPARISON

Table 20: The G1000 and Avidyne Entegra systems record similar data parameters in their logs but sometimes differ in their parameter names or units used.

Parameter	G1000 Parameter ID	G1000 Units	Avidyne Parameter ID	Avidyne Units
Time Stamp			timeStamp	
Local Date	Lcl Date	mm/dd/yyyy	mUtcDate	mm:dd:yyyy
Local Time	Lcl Time	hh:mm:ss	mUtcTime	hh:mm:ss
Timezone	UTCOfst	hh:mm		
Time in Service			minutesInService	minutes
Active Waypoint Identifier	AtvWpt	ident	mNxWptID	
Distance to Next Waypoint	WptDst	nm	DistanceToWpt	nm
Bearing to Next Waypoint	WptBrg	degrees	ActiveBearing	degrees
Estimated Time En Route			mEteInSeconds	seconds
Latitude	Latitude	degrees	mLatitude	degrees
Longitude	Longitude	degrees	mLongitude	degrees
Altitude	AltB	feet Baro	altitude; baroCorrectedAlt	feet
Altitude Valid			altitudeValid; baroCorrectedAltValid	
Altitude Bug			AltBug	feet
Barometer Setting	BaroA	inches	baroSetting	inHg
Barometer Setting Valid			baroSettingValid	
Barometer Bug			mBaroBug	inHg
MSL Altitude	AltMSL	feet MSL		
Density Altitude			densityAltitude	feet
Density Altitude Valid			densityAltitudeValid	
Outside Air Temperature	OAT	degrees C	totalTemperature	degrees C
Total Temperature Valid			totalTemperatureValid	
Indicated Airspeed	IAS	kt	indicatedAirspeed	kt
Indicated Airspeed Bug			mIasBug	kt

Parameter	G1000 Parameter ID	G1000 Units	Avidyne Parameter ID	Avidyne Units
Indicated Airspeed Valid			indicatedAirspeedValid	
Ground Speed	GndSpd	kt	mGroundSpeed	kt
Vertical Speed	VSpd	fpm	altitudeRate	fpm
Altitude Rate Valid			altitudeRateValid	
Vertical Speed Indicator Bug			mVsiBug	fpm
Pitch	Pitch	degrees	pitch	
Pitch Valid			pitchValid	
Pitch Rate			Pitch Rate	degrees/s
Pitch Rate Valid			Pitch Rate Valid	
Roll	Roll	degrees	roll	
Roll Valid			rollValid	
Roll Rate			Roll Rate	degrees/s
Roll Rate Valid			Roll Rate Valid	
Yaw Rate			Yaw rate	degrees/s
Yaw Rate Valid			Yaw Rate Valid	
Turn Rate			rateofTurn	
Turn Rate Valid			rateofTurnValid	
Lateral Acceleration	LatAc	G	lateralAcceleration; Lat Accel	m/s^2
Lateral Acceleration Valid			lateralAcceleration Valid; Lat Accel Valid	
Vertical Acceleration	NormAc	G	Norm Accel	m/s^2
Vertical Acceleration Valid			Norm Accel Valid	
Longitudinal Acceleration			Long Accel	m/s^2
Longitudinal Acceleration Valid			Long Accel Valid	
Heading	HDG	degrees	magHeading	
Heading Bug			mHdgBug	degrees
Magnetic Heading Valid			magHeadingValid	
Track	TRK	degrees	mGroundTrack	degrees
Voltage 1	volt1; volt2	volts		
Amperage 1	amp1; amp2	amps		
Fuel Flow	E1 FFlow	gph	fuelflowL; fuelflowR; fuelFlowL; fuelFlowR	gph; lbph

Parameter	G1000 Parameter ID	G1000 Units	Avidyne Parameter ID	Avidyne Units
Oil Temperature	E1 OilT	degrees F	oilTempL / oilTempR	degrees F
Oil Pressure	E1 OilP	psi	oilPresL / oilPresR / oilPressL / oilPressR	psi
Manifold Absolute Pressure	E1 MAP	Hg	manPresL; manPresR	inHg
Engine Rotations per Minute	E1 RPM	rpm	tachL; tachR	rpm
Engine Percent Power			percentPowerL; percentPowerR	%
Engine Percent Torque			engineTorquePercentL; engineTorquePercentR	%
Turbine Rotations per Minute			engineNgPercentL; engineNgPercentR	%
Propeller Rotations per Minute			engineNpPercentL; engineNpPercentR	%
Inlet Turbine Temperature			ittDegCL; ittDegCR	
Cylinder Head Temperature	E1 CHT1; E1 CHT2; E1 CHT 3; E1 CHT4; E1 CHT5; E1 CHT6	degrees F		
Exhaust Gas Temperature	E1 EGT1; E1 EGT2; E1 EGT 3; E1 EGT4; E1 EGT5; E1 EGT6	degrees F		
Cool Temperature			coolTempL; coolTempR	degrees F
Altitude GPS	AltGPS	ft wgs		
True Airspeed	TAS	kt	trueAirspeed	kt
True Airspeed Valid			trueAirspeedValid	
Airspeed Trend			airspeedTrend	
Airspeed Trend Valid			airspeedTrendValid	
	HSIS	enum		
Course	CRS	degrees	ActiveCourse	degrees
Desired Course			mDtkOrBrg; DesiredCourse	degrees

Parameter	G1000 Parameter ID	G1000 Units	Avidyne Parameter ID	Avidyne Units
Navigational Frequency	NAV1; NAV2	MHz	VhfFreq	
Primary Navigation Source			ucPriNavSource	
Communication Frequency	COM1; COM2	MHZ		
Horizontal Course Deviation Indicator	HCDI	fsd	HdiDeviation	%
Horizontal Course Deviation Indicator Source			HdiSource	
Vertical Course Deviation Indicator	VCDI	fsd	VdiDeviation	%
Vertical Course Deviation Indicator Source			VdiSource	
Wind Speed	WndSpd	kt		
Wind Direction	WndDr	degrees		
Magnetic Variation	MagVar	degrees		
Automatic Flight Control System On	AfcsOn	bool		
*	RollM	enum		
*	PitchM	enum		
Roll	RollC	degrees	fdRoll	
Pitch	PitchC	degrees	fdPitch	
GPS Vertical Speed	VSpdG	fpm		
GPS Fix	GPSfix	enum	GpsHold	
Horizontal Alert Limit	HAL	mt	HdiDeviationLimit	
Vertical Alert Limit	VAL	mt	VdiDeviationLimit	
*	HPLwas	mt		
*	HPLfd	mt		
*	VPLwas	mt		
Active Annunciators			apAnnunciators	
Logic States			logicStates	
Map Format			mMapFormat	enum
Map Range			mMapRangeIndex	
Flags			Flags; FlagsL; FlagsR; WaasFlags	
Saturated			saturated	
Saturated Valid			saturatedValid	

Parameter	G1000 Parameter ID	G1000 Units	Avidyne Parameter ID	Avidyne Units
Go/No-go			GoNogo; mpuNoGo; iruNoGo; magNoGo	
Needle Text Type			mNeedleTextType	enum
Dh Alert			mDhAlert	
Synthetic Rate Alarm			SyntheticRateAlarm	
Longterm Bias Drift Alarm			LongtermBiasDrift Alarm	
Bias Cutout Alarm			BiasCutoutAlarm	

APPENDIX B. SURVEY

Data-driven safety feedback as part of debrief for General Aviation pilots

Start of Block: Informed Consent

Dear Aviation Colleague,

My name is Nicoletta Fala, and I am a Ph.D. candidate working with Prof. Karen Marais at the School of Aeronautics and Astronautics at Purdue University. We are seeking your input on post-flight debrief feedback in this survey.

The motivation behind this research is the unacceptably high number of general aviation accidents. Our overall goal is to use flight data of various sources to help improve general aviation safety. We are trying to understand how different kinds of safety feedback affect risk perception among general aviation pilots.

During the survey, you will be asked to review flight data from four flights and answer specific questions on the safety of each flight. We will then ask you a few demographic questions. The survey should take approximately 20 minutes to complete. During the survey, you will not be able to go back to the previous flight safety questions. You will, however, have the opportunity to review and change the demographic questions as you wish. You may choose to not answer some questions and you may stop the survey at any time without any repercussion to you. If you do not wish to complete the survey in one sitting, you may save your progress and return where you left off if you use the same computer to re-access the link. No personally identifiable information is being asked, analyzed or reported. All responses will be anonymous and in aggregate at the end of the study.

Your participation in this survey is voluntary. You must be at least 18 years old to participate in this research. Thank you for your time and your cooperation. Your responses are greatly appreciated and will hopefully enable the general aviation community to improve their safety record. If you have any questions regarding the survey or the information contained within, please feel free to contact the researchers directly either at nfala@purdue.edu or kmarais@purdue.edu.

Page Break

RESEARCH PARTICIPANT CONSENT FORM

Data-driven safety feedback as part of debrief for General Aviation pilots

Principal Investigator: Associate Professor Karen Marais

School of Aeronautics and Astronautics

Purdue University

IRB Protocol # 1804020499

What is the purpose of this study?

This study seeks to evaluate whether data-driven post-flight debrief can be used to impact how pilots react to safety information. As a pilot, you can help us answer our research questions by evaluating the risk of hypothetical flights that you will have the chance to review. Through this research, we hope to come up with recommendations on how to communicate risk to pilots in a flight debrief format.

What will I do if I choose to be in this study?

If you choose to participate in this survey, you will be asked to review sample debrief screens of hypothetical flights. The screens will help you visualize the flight and give you information regarding the takeoff phase of each flight. At the end of the survey, we will also ask you some demographic questions.

How long will I be in the study?

This survey should take you approximately 20 minutes to complete.

What are the possible risks or discomforts?

The risk level to participating in this study is minimal, no greater than you would encounter in daily life or during the performance of routine psychological exams or tests. Breach of confidentiality is a possible risk, however no identifiable information will be collected during the study.

Are there any potential benefits?

There are no direct benefits to participating in this study. We believe you will enjoy debriefing these flights. In the future, the results of this study may help us make General Aviation safer by understanding how to communicate risk better.

Will information about me and my participation be kept confidential?

All demographic information and answers to questions are anonymous. We will not be asking for or collecting any identifiable information in this survey. All demographic information and answers to questions will be kept indefinitely on a hard drive located in Armstrong Hall, for use in future research and academic publications.
The project's research records may be reviewed by departments at Purdue University responsible for regulatory and research oversight.

What are my rights if I take part in this study?

Your participation in this study is voluntary. You may choose not to participate, or, if you agree to participate, you can withdraw your participation at any time without penalty or loss of benefits to which you are otherwise entitled. If you decide to stop the survey without finishing, some of your responses may still be usable to the researchers.

Who can I contact if I have questions about the study?

If you have questions, comments, or concerns about this research project, you can talk to one of the researchers. Please contact Prof. Karen Marais at (765) 494-0063 or kmarais@purdue.edu. If you have questions about your rights while taking part in the study or have concerns about the treatment of research participants, please call the Human Research Protection Program at (765) 494-5942, email (irb@purdue.edu) or write to:

Human Research Protection Program - Purdue University
Ernest C. Young Hall, Room 1032
155 S. Grant St.,
West Lafayette, IN 47907-2114

Do you consent to participate in this research project?

○ Yes (1)

○ No (2)

End of Block: Informed Consent

Start of Block: 0.0

During this survey, you will be presented with three sets of semi-interactive debrief screens for the takeoff phase of flight for a Cessna 172. The first set of screens is a tutorial so that you can get comfortable with navigating through the different screens. After completing the tutorial, you will have a chance to debrief and answer questions on three flights.

You can obtain more information on the performance of a Cessna 172 here. You can refer back to this document as you go through the different screens. During the survey, you will have as much time as you need to review the debrief screens, but once you click on "Next" to proceed to the questions, you will not be able to return to the debrief.

When you are ready to take the actual survey, proceed to the next screen. Please remember that you will not be able to return to the debrief after clicking "Next."

<<Interactive tutorial version of debrief tool.>>

Thank you for completing the tutorial; you can now move on to reviewing and evaluating takeoffs.

End of Block: 0.0

Note: The following blocks are repeated three times for three different flights, before moving to the demographics block. The questions displayed take the form of the *Safety* block or the *Risk* block, both of which are included here. Only one questions block is displayed for each flight reviewed.

Start of Block: 1.1-1.4

Review the following takeoff phase of flight as presented in these debrief screens, taking as much time as you need. The aircraft involved is a Cessna 172.

The debrief screens are semi-interactive: Under "Segments Manager," click on "takeoff KOSU RWY 27L" to choose the takeoff segment. Then click on each event you want to further investigate from the "takeoff safety information" list on the right.

When you are ready to answer questions about this takeoff, proceed to the next screen. Note that you will not be able to return to the debrief after clicking "Next."

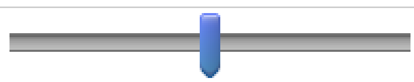
<<Randomized interactive debrief tool.>>

End of Block: 1.1-1.4

Start of Block: Questions [Safety]

Given the information presented to you, how safe would you say this takeoff was?

	Not safe at all			Extremely safe	
	1	2	3	4	5



In this takeoff, which of the following would concern you, if any?

- ☐ Centerline deviation (1)
- ☐ Rotation airspeed (2)
- ☐ Engine RPM (3)
- ☐ Takeoff distance (4)
- ☐ Wind (5)

QS3 Optional comments

____

____

____

____

____

Page Break

What changes (up to 5) do you think you could make to an upcoming flight after the information presented here, if any?

- ◯ Change 1 (1) ____
- ◯ Change 2 (2) ____
- ◯ Change 3 (3) ____
- ◯ Change 4 (4) ____
- ◯ Change 5 (5) ____

Page Break

Display This Question:

If What changes (up to 5) do you think you could make to an upcoming flight after the information pr... Change 1 Is Not Empty

How likely are you to make each of these changes to an upcoming flight?

	Not likely at all			Extremely likely	
	1	2	3	4	5
${QS3/ChoiceTextEntryValue/1} ()					
${QS3/ChoiceTextEntryValue/2} ()					
${QS3/ChoiceTextEntryValue/3} ()					
${QS3/ChoiceTextEntryValue/4} ()					
${QS3/ChoiceTextEntryValue/5} ()					

Display This Question:

If What changes (up to 5) do you think you could make to an upcoming flight after the information pr... Change 1 Is Not Empty

How important do you think each of these changes is to improving safety on takeoff?

	Not important at all			Extremely important	
	1	2	3	4	5
${QS3/ChoiceTextEntryValue/1} ()					
${QS3/ChoiceTextEntryValue/2} ()					
${QS3/ChoiceTextEntryValue/3} ()					
${QS3/ChoiceTextEntryValue/4} ()					
${QS3/ChoiceTextEntryValue/5} ()					

End of Block: Questions [Safety]

Start of Block: Questions [Risk]

Given the information presented to you, how risky would you say this takeoff was?

	Not risky at all			Extremely risky	
	1	2	3	4	5

In this takeoff, which of the following would concern you, if any?

- ☐ Centerline deviation (1)
- ☐ Rotation airspeed (2)
- ☐ Engine RPM (3)
- ☐ Takeoff distance (4)
- ☐ Wind (5)

QS3 Optional comments

____

____

____

____

____

Page Break

What changes (up to 5) do you think you could make to an upcoming flight after the information presented here, if any?

- ◯ Change 1 (1) ____
- ◯ Change 2 (2) ____
- ◯ Change 3 (3) ____
- ◯ Change 4 (4) ____
- ◯ Change 5 (5) ____

Page Break

Display This Question:

If What changes (up to 5) do you think you could make to an upcoming flight after the information pr... Change 1 Is Not Empty

How likely are you to make each of these changes to an upcoming flight?

	Not likely at all				Extremely likely
	1	2	3	4	5
${QS3/ChoiceTextEntryValue/1} ()					
${QS3/ChoiceTextEntryValue/2} ()					
${QS3/ChoiceTextEntryValue/3} ()					
${QS3/ChoiceTextEntryValue/4} ()					
${QS3/ChoiceTextEntryValue/5} ()					

Display This Question:

If What changes (up to 5) do you think you could make to an upcoming flight after the information pr... Change 1 Is Not Empty

How important do you think each of these changes is to reducing risk on takeoff?

	Not important at all			Extremely important	
	1	2	3	4	5
${QS3/ChoiceTextEntryValue/1} ()					
${QS3/ChoiceTextEntryValue/2} ()					
${QS3/ChoiceTextEntryValue/3} ()					
${QS3/ChoiceTextEntryValue/4} ()					
${QS3/ChoiceTextEntryValue/5} ()					

End of Block: Questions [Risk]

Start of Block: Demographics

You are almost done! We will next ask you some quick demographic questions that will help us improve the quality of our analysis.

Page Break

QD1 How old are you?

- ○ 18-24 (1)
- ○ 25-34 (2)
- ○ 35-44 (3)
- ○ 45-54 (4)
- ○ 55-64 (5)
- ○ 65 or older (6)

QD2 What gender do you identify with?

- ○ Male (1)
- ○ Female (2)
- ○ Other (3) ____
- ○ Do not wish to specify (4)

QD3 What is the highest level of education you have completed?

- Some high school (1)
- High school graduate or equivalent (2)
- Some college (3)
- 2-year degree (11)
- 4-year degree (12)
- Master's degree (13)
- Doctorate or Professional degree (14)

QD4 What is your occupation?

____

QD5 What kind of pilot's license do you currently have?

- No certificate (1)
- Student (2)
- Sport (3)
- Recreational (4)
- Private (5)
- Commercial (6)
- Airline Transport (7)

QD6 Which ratings or endorsements do you currently have?

- ☐ Single-engine (1)
- ☐ Multi-engine (2)
- ☐ Instrument (3)
- ☐ Rotorcraft-Helicopter (4)
- ☐ Glider (5)
- ☐ Lighter-than-air (6)
- ☐ Seaplane (7)
- ☐ Tailwheel (8)
- ☐ High altitude (9)
- ☐ High performance (10)
- ☐ Flight instructor (11)
- ☐ Instrument flight instructor (12)
- ☐ Multi-engine flight instructor (13)

QD7 How many years of flying experience do you have? Round to the nearest year.

____

QD8 Was your flight training under Part 61 or Part 141?

- ○ Part 61 (1)
- ○ Part 141 (2)
- ○ Combination/both (3)
- ○ I do not know (4)

QD9 What kind of avionics do you most frequently use in your flying?

- ○ Mostly steam gauges (1)
- ○ Mostly glass cockpit (2)
- ○ I fly both equally frequently (3)

QD10 How many flight hours do you have logged (approximately)?

____

QD11 How often do you fly?

- ○ Once a week (1)
- ○ 2-7 days a week (2)
- ○ Once a month (3)
- ○ Once every few months (4)
- ○ Rarely/never (5)

QD12 How often do you participate in aviation safety programs and seminars (such as WINGS, FAASTeam seminars, AOPA training videos, etc.)?

- Monthly (1)
- 2-3 times a year (2)
- Once a year (3)
- Once every two years (4)
- Never (5)

QD13 Where is your home airport base? (ICAO identifier or city/state)

____

QD14 Have you used commercial debrief products, like CloudAhoy, before?

- Yes (1)
- No (2)

End of Block: Demographics

APPENDIX C. SURVEY RESPONSES

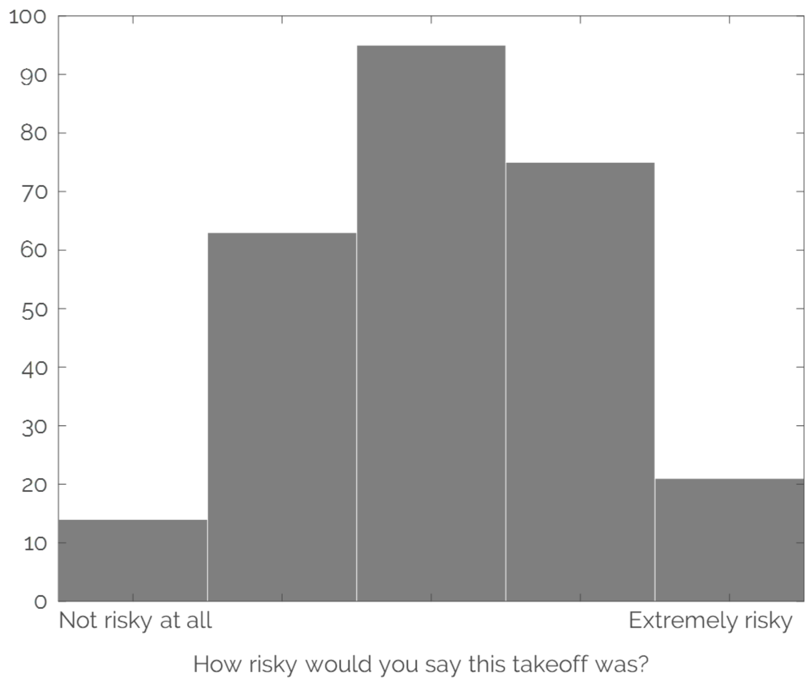


Figure 24: The respondents answered the question "How risky would you say this takeoff was?" using a 5-point Likert scale. For Flight A, most respondents opted for a neutral value around the center, but ~13% used the extreme values of *not risky at all* and *extremely risky*.

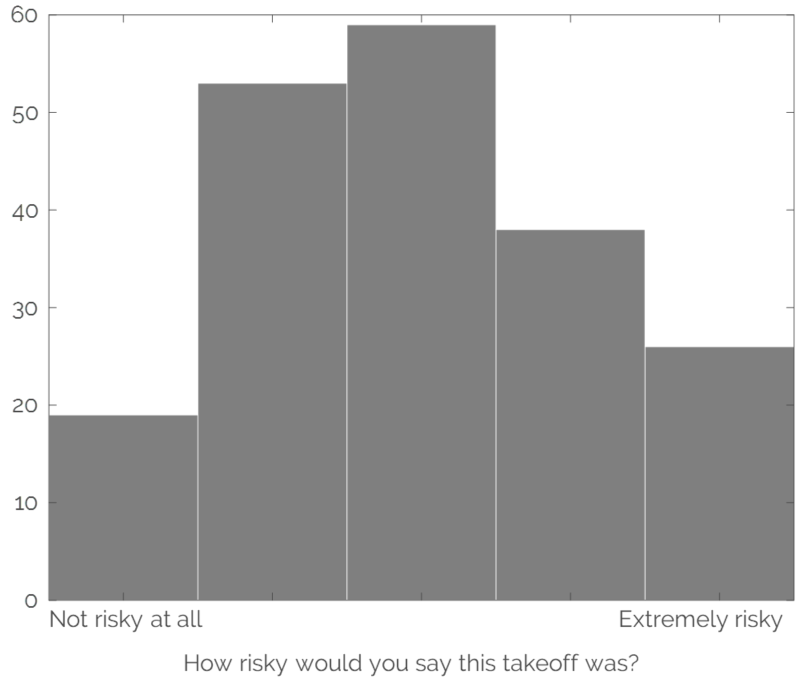


Figure 25: The respondents in Flight B also congregated around the mid-point, however, many more of them (23%) chose the extreme values of *not risky at all* and *extremely risky*. The overall number of responses decreased for Flight B compared to Flight A.

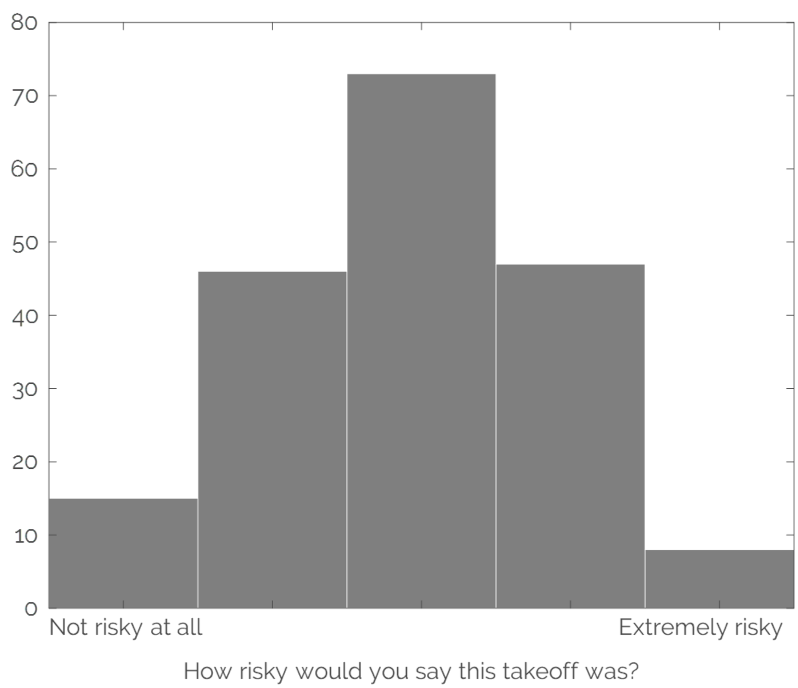


Figure 26: Most respondents for Flight C also opted for a neutral value around the center, with only ~12% using the extreme values of *not risky at all* and *extremely risky*.

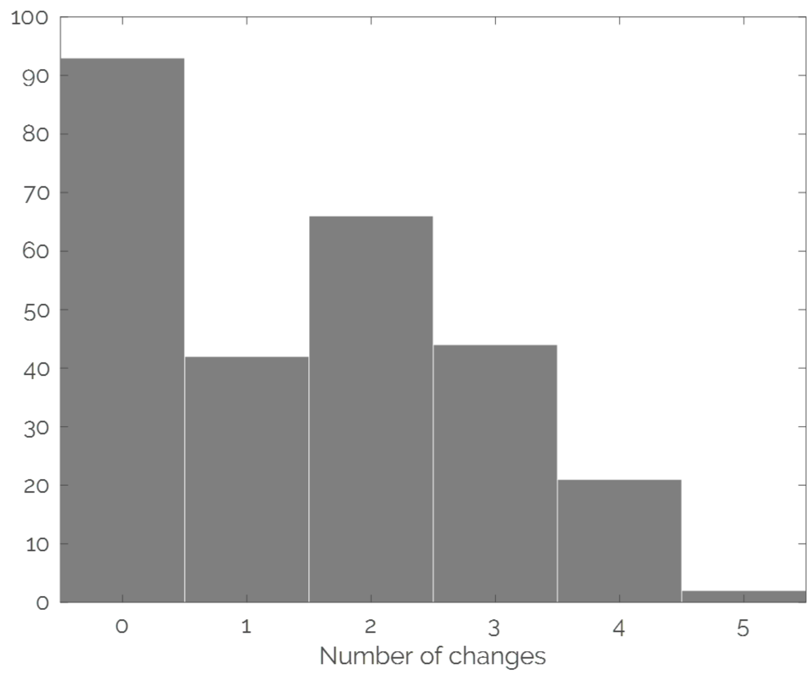


Figure 27: Most respondents said they would make two changes to their flying after reviewing their debrief for Flight A. Most of these changes referred to the centerline and the airspeed—pilots said they would maintain better rudder control to keep the nose on the centerline, and be more patient in waiting until 55 knots to rotate. Some respondents also referred to the wind, takeoff distance, and engine RPM, with suggestions to use a different runway, avoid the intersection departure, talk to a mechanic to evaluate the engine performance, and potentially abort the takeoff because of centerline deviation.

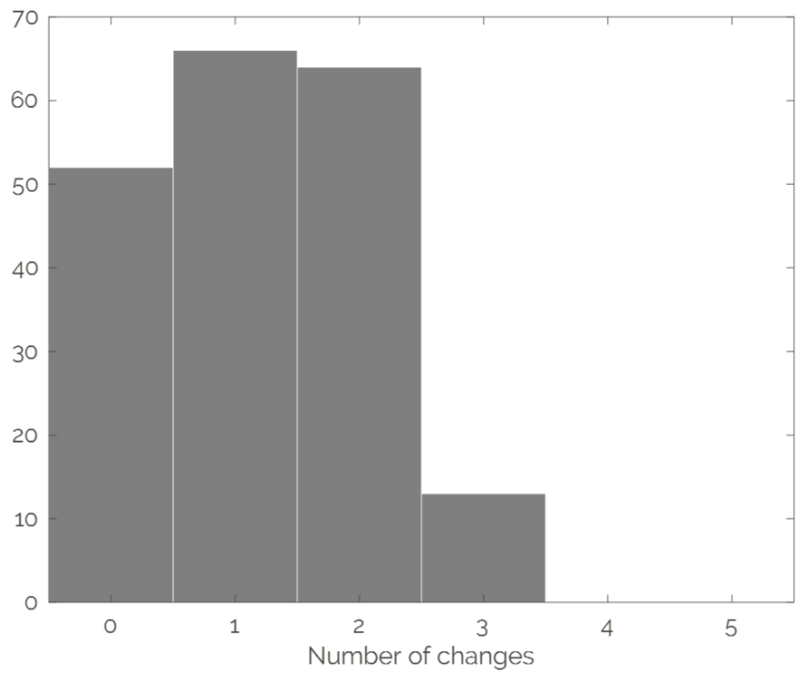


Figure 28: Respondents in Flight B said they would make up to three changes at most. Most pilots only wrote in one or two changes. Since Flight B had less hazardous states than Flight A, the discrepancy in the number of changes is reasonable. The changes in Flight B focused on engine issues and rotation airspeed.

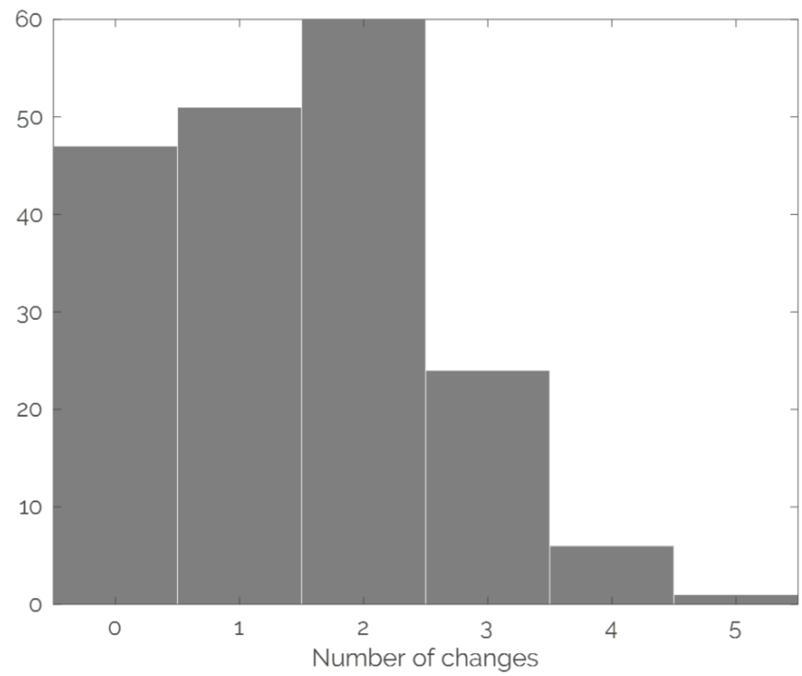


Figure 29: Changes suggested for Flight C included better directional control, choosing a different runway, and rotating at a higher airspeed. Most respondents said they would make up to two changes to their flying.

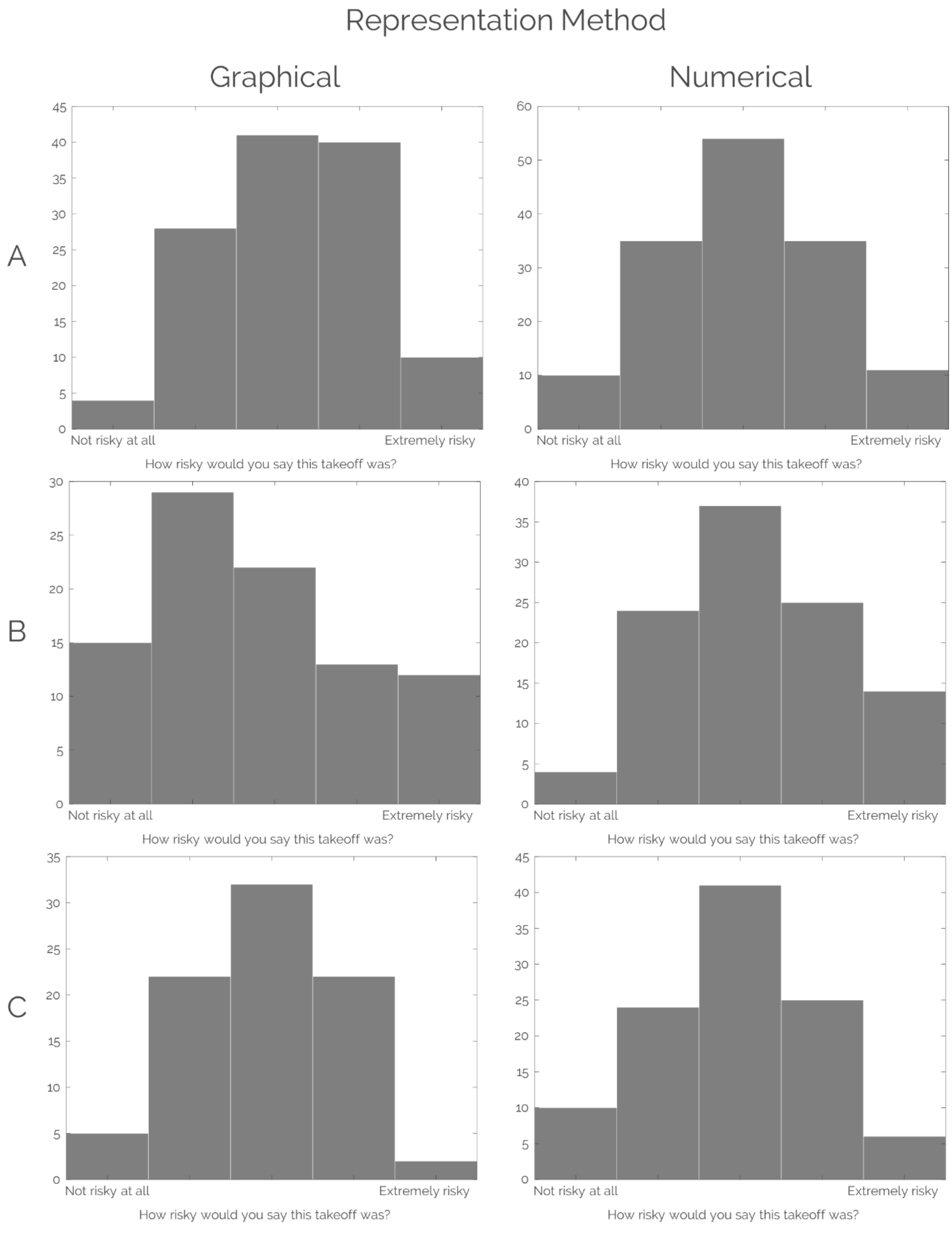


Figure 30: Survey respondents could see their flight debrief information *graphically* (left) or *numerically* (right). The three rows correspond to the three flights the respondents reviewed.

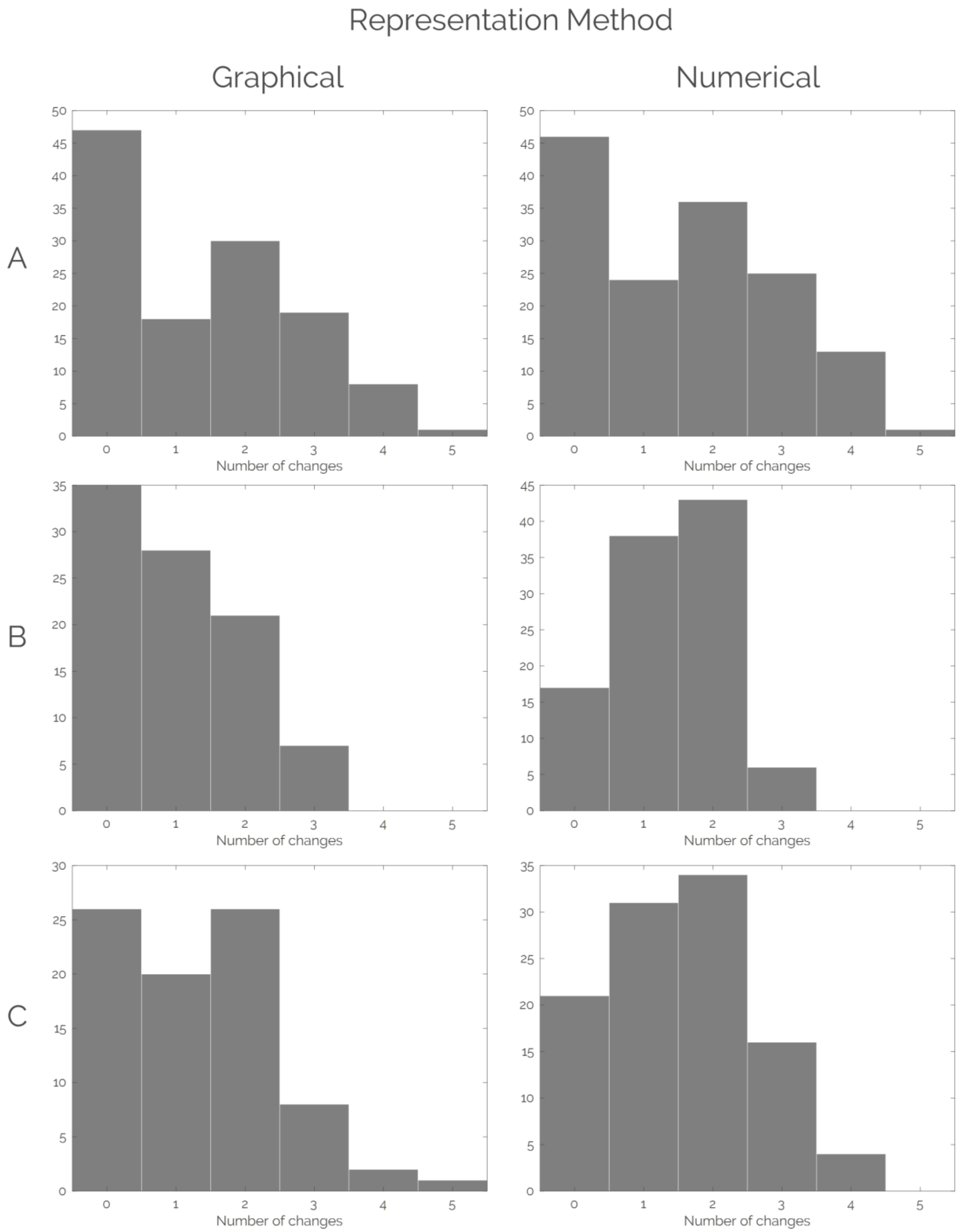


Figure 31: Respondents were more likely to do nothing to change the behaviors they identified when I presented their debrief to them graphically (left). The frequency of zero changes goes down in the numerical case (right).

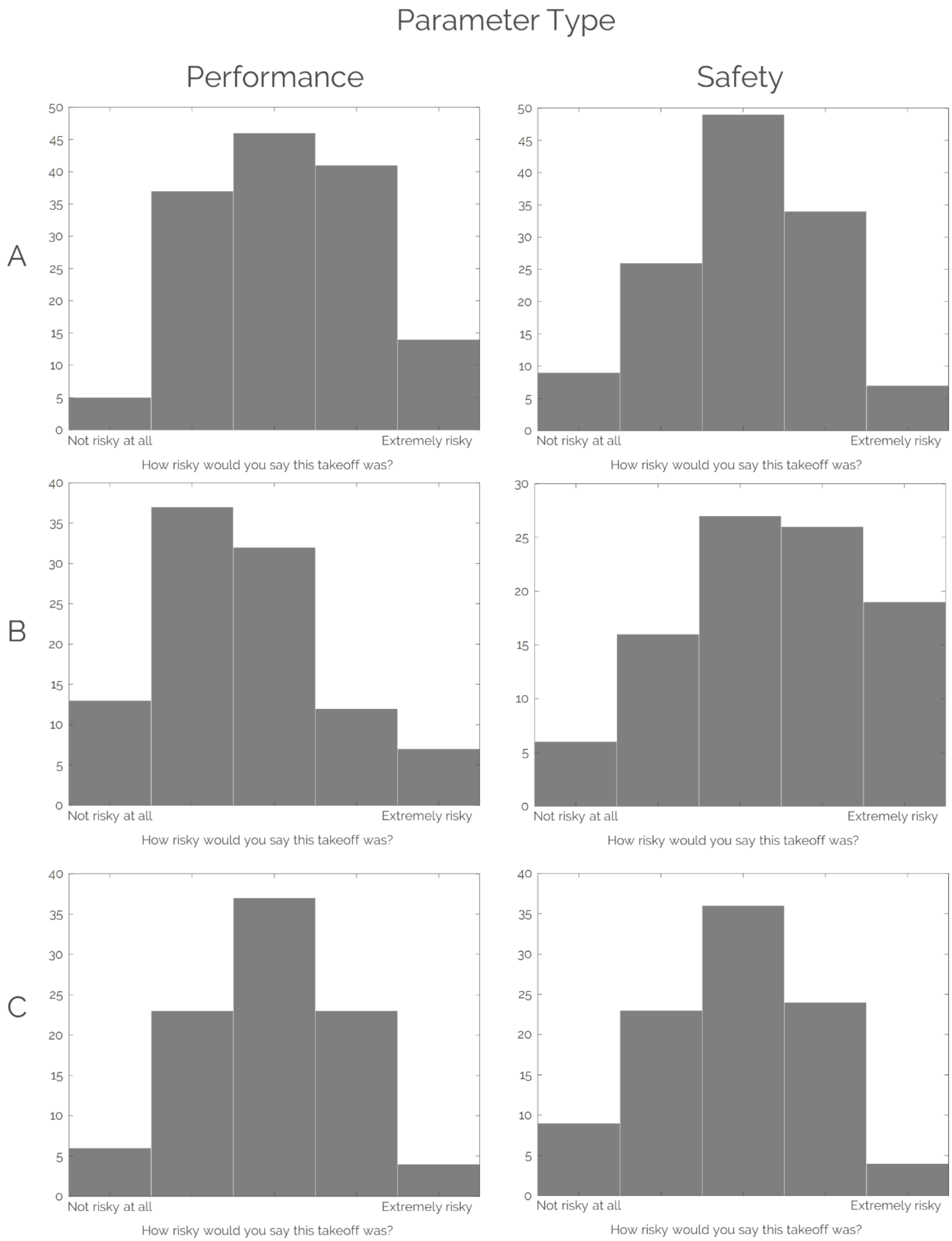


Figure 32: The survey communicated information in terms of *performance parameters* (left) and *safety parameters* (right).

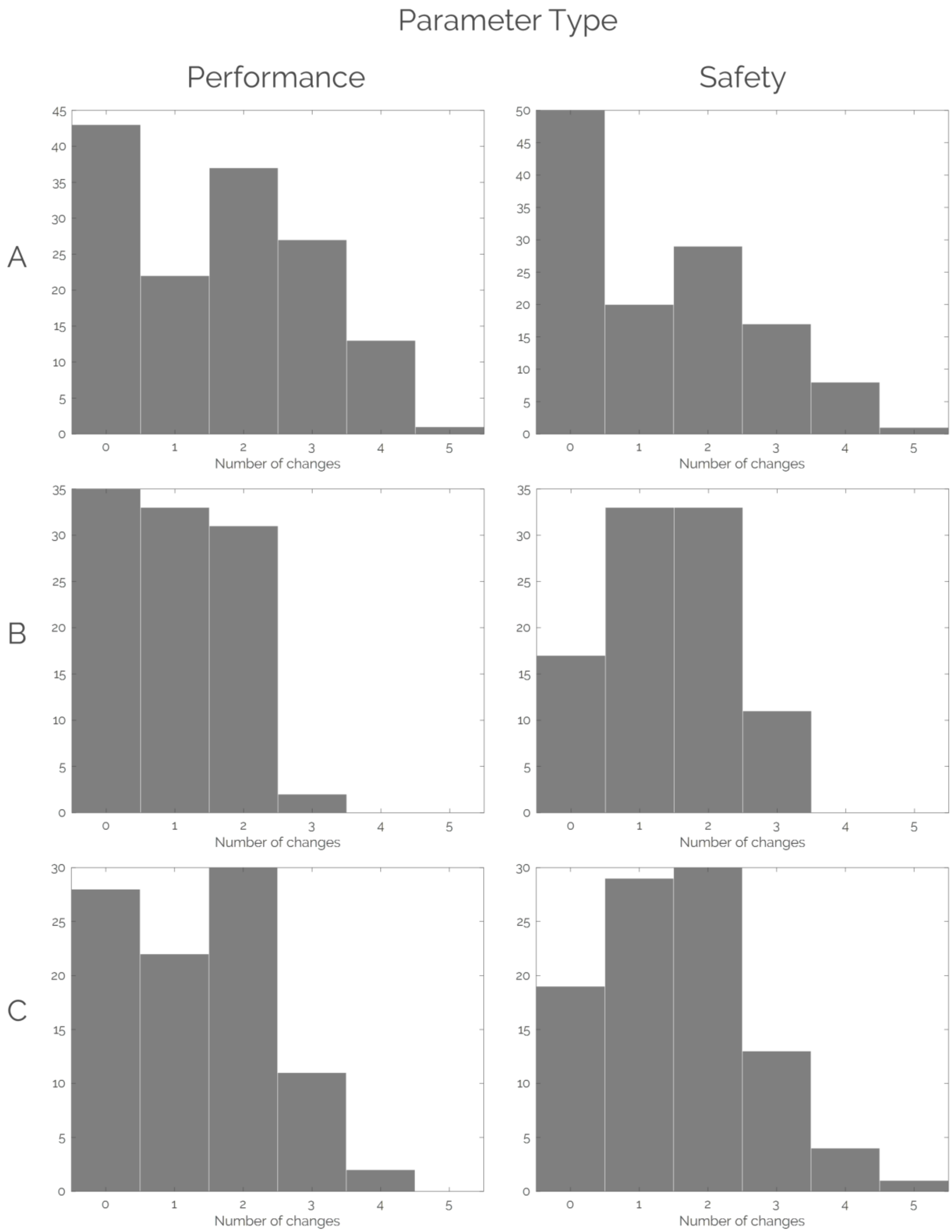


Figure 33: The safety parameter version of the debrief reduced the respondents who opted to continue without making any changes in Flight B and C, but increased the "no changes" responses in Flight A.

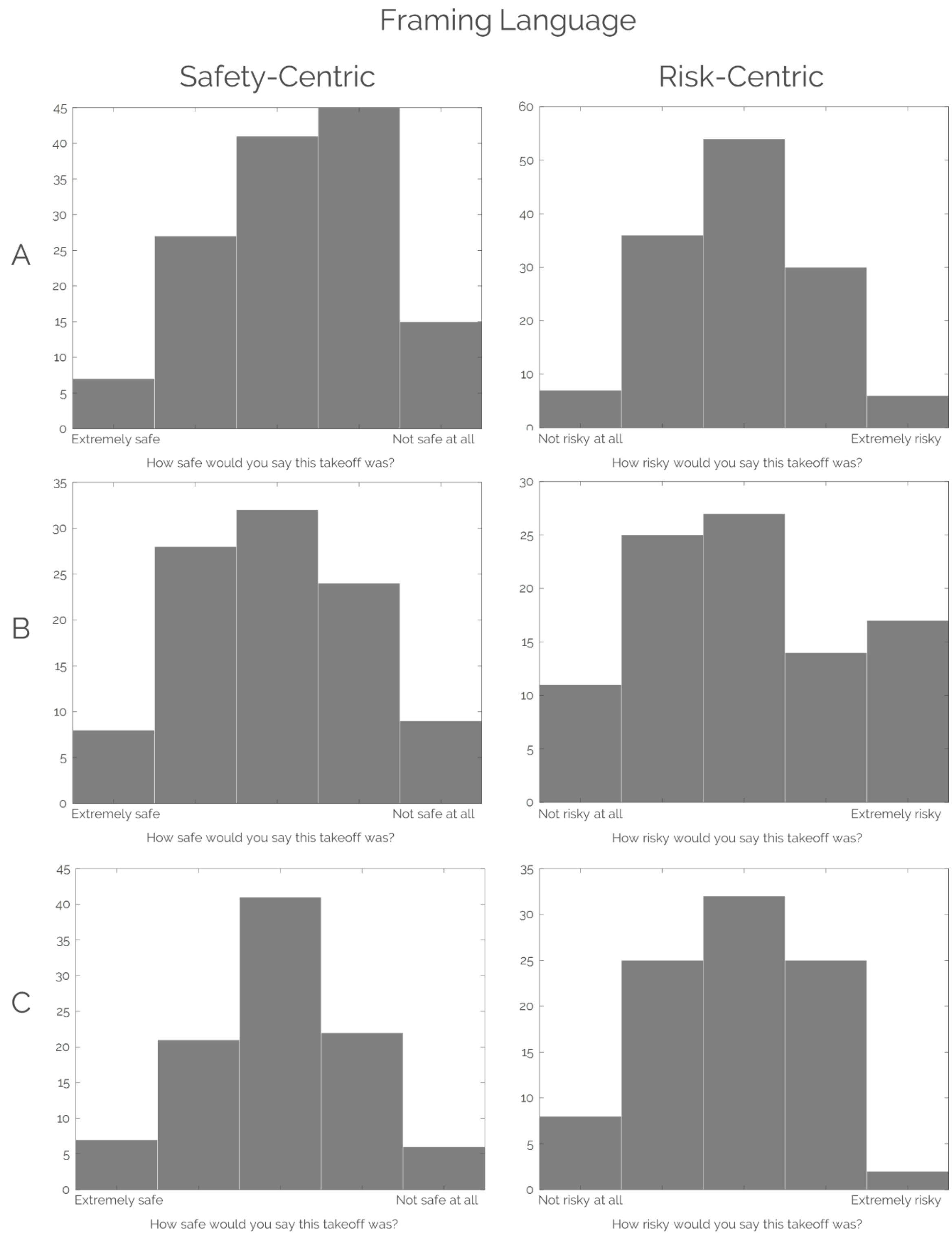


Figure 34: The survey asked pilots to rate the *risk* or *safety* of the takeoff in the flight they debriefed. The Likert-scale was inverted in each case to maintain consistency.

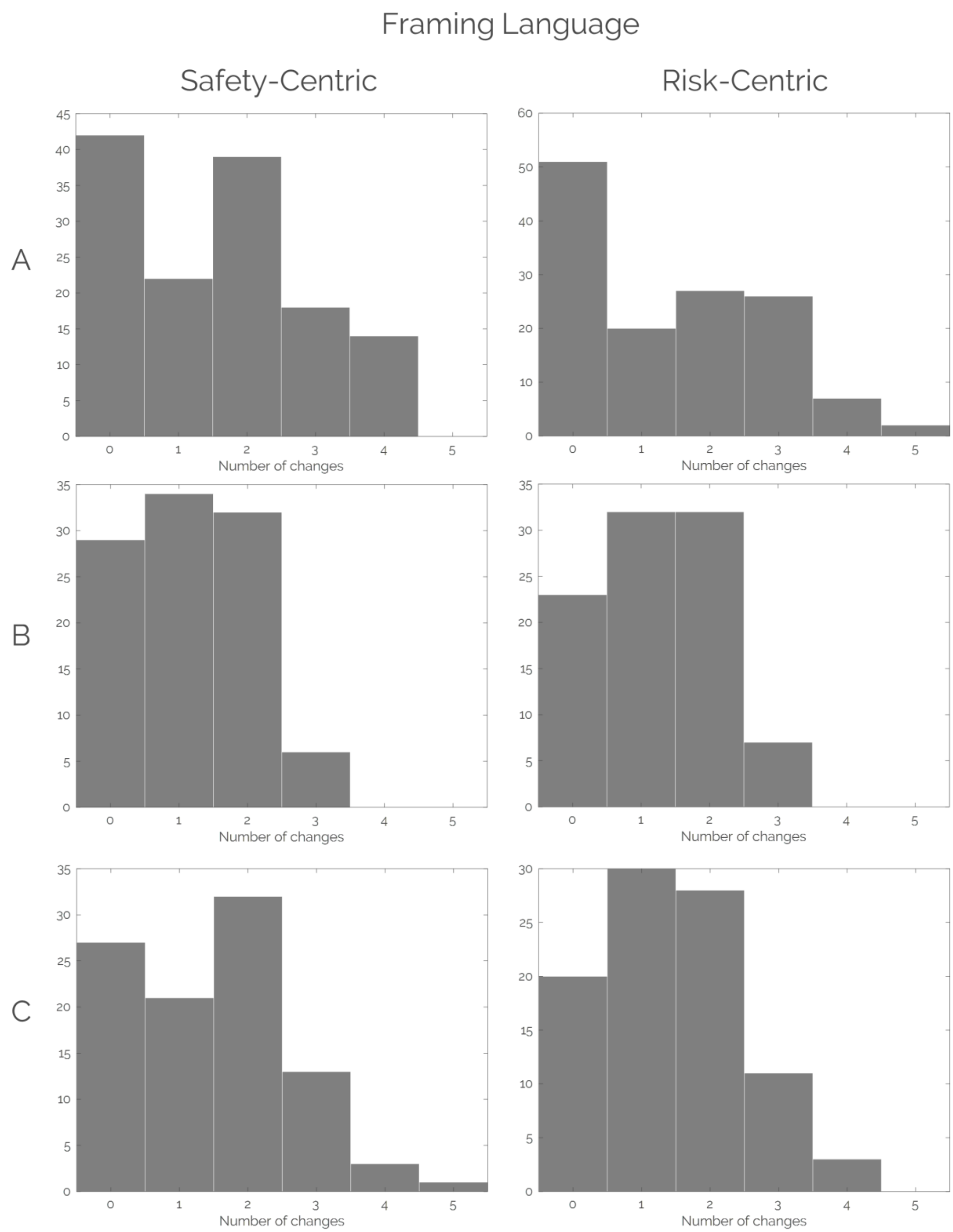


Figure 35: The framing language did not consistently affect the number of changes pilots suggested in any of the three flights.

Table 21: The Scheirer-Ray-Hare test for the risk perception in Flight A identified framing language as a main effect and the representation method and parameter type factors as an interaction effect.

y ~ parameter + language				
	df	**Sum Sq.**	**H**	**p-value**
parameter	1	3426	0.6198	0.43111
language	1	35420	6.4089	0.01135
parameter:language	1	747	0.1351	0.71316
Residuals	264	1436035		
y ~ representation + language				
	df	**Sum Sq.**	**H**	**p-value**
representation	1	11623	2.1031	0.14700
language	1	40009	7.2392	0.00713
representation:language	1	1395	0.2523	0.61543
Residuals	264	1422601		
y ~ representation + parameter				
representation	1	11623	2.1031	0.14700
parameter	1	3356	0.6073	0.43582
representation:parameter	1	31252	5.6548	0.01741
Residuals	264	1429396		

Table 22: The Scheirer-Ray-Hare test for the number of changes pilots suggested after debriefing their flights in Flight A did not identify any main or interaction effects.

y ~ parameter + language				
	df	**Sum Sq.**	**H**	**p-value**
parameter	1	20559	3.6619	0.05567
language	1	2315	0.4124	0.52077
parameter:language	1	38	0.0068	0.93425
Residuals	264	1476124		
y ~ representation + language				
	df	**Sum Sq.**	**H**	**p-value**
representation	1	6476	1.15353	0.28281
language	1	3379	0.60182	0.43789
representation:language	1	5004	0.89123	0.34514
Residuals	264	1484178		
y ~ representation + parameter				
representation	1	6476	1.1535	0.282811
parameter	1	20688	3.6848	0.054911
representation:parameter	1	6257	1.1145	0.291097
Residuals	264	1465615		

Table 23: The Scheirer-Ray-Hare test for the risk perception in Flight B identified parameter type and representation methods as main effects but no interaction effects.

y ~ parameter + language				
	df	**Sum Sq.**	**H**	**p-value**
parameter	1	60657	20.2269	0.00001
language	1	3	0.0011	0.97337
parameter:language	1	1211	0.4038	0.52515
Residuals	191	519904		
y ~ representation + language				
	df	**Sum Sq.**	**H**	**p-value**
representation	1	22436	7.4815	0.00623
language	1	106	0.0354	0.85082
representation:language	1	274	0.0913	0.76259
Residuals	191	488855		
y ~ representation + parameter				
representation	1	22436	7.4815	0.006233
parameter	1	59368	19.7970	0.000009
representation:parameter	1	11117	3.7070	0.054186
Residuals	191	488855		

Table 24: The Scheirer-Ray-Hare test for the number of changes pilots suggested after debriefing their flights in Flight B did not identify any interaction effects, but did identify parameter and representation as significant main effects.

y ~ parameter + language				
	df	**Sum Sq.**	**H**	**p-value**
parameter	1	25102	8.6928	0.00319
language	1	1356	0.4695	0.49321
parameter:language	1	1568	0.5429	0.46121
Residuals	191	532177		
y ~ representation + language				
	df	**Sum Sq.**	**H**	**p-value**
representation	1	25524	8.8391	0.00295
language	1	698	0.2419	0.62284
representation:language	1	2196	0.7606	0.38313
Residuals	191	531784		
y ~ representation + parameter				
representation	1	25524	8.8391	0.002948
parameter	1	20688	8.3857	0.003782
representation:parameter	1	6257	1.5725	0.209843
Residuals	191	505923		

Table 25: The Scheirer-Ray-Hare test for the risk perception in Flight C identified no main or interaction effects.

y ~ parameter + language				
	df	**Sum Sq.**	**H**	**p-value**
parameter	1	214	0.07834	0.77956
language	1	1411	0.51687	0.47218
parameter:language	1	7110	2.60493	0.10653
Residuals	185	504366		
y ~ representation + language				
	df	**Sum Sq.**	**H**	**p-value**
representation	1	5	0.00177	0.96643
language	1	1392	0.50986	0.47520
representation:language	1	3995	1.46383	0.22632
Residuals	185	507709		
y ~ representation + parameter				
representation	1	5	0.00177	0.96643
parameter	1	214	0.07842	0.77945
representation:parameter	1	3638	1.33310	0.24825
Residuals	185	509243		

Table 26: The Scheirer-Ray-Hare test for the number of changes pilots suggested after debriefing their flights in Flight C did not identify any main or interaction effects.

y ~ parameter + language				
	df	**Sum Sq.**	**H**	**p-value**
parameter	1	4392	1.57665	0.20924
language	1	23	0.00826	0.92758
parameter:language	1	84	0.03025	0.86192
Residuals	185	519230		
y ~ representation + language				
	df	**Sum Sq.**	**H**	**p-value**
representation	1	6172	2.21543	0.13664
language	1	698	0.00834	0.92724
representation:language	1	2196	0.07771	0.78042
Residuals	185	531784		
y ~ representation + parameter				
representation	1	6172	2.2154	0.13664
parameter	1	4357	1.5640	0.21108
representation:parameter	1	5966	2.1417	0.14334
Residuals	185	507234		

Table 27: The results of the ANOVA on the risk perception responses from Flight A indicate that the *Framing Language* factor likely moved the location of the mean response. The *Representation Method* and *Parameter Type* factors combined may influence the results. Rows shaded in the darker gray correspond to parameters that are significant at the 0.05 significance level, with the lighter gray color used to identify rows that came close to the 0.05 significance level.

Source	**Sum Sq.**	**d.f.**	**Mean Sq.**	**F**	**Prob>F**
Repres	3.302	1	3.30242	3.34	0.0689
Param	0.237	1	0.23702	0.24	0.6249
Lang	6.845	1	6.84477	6.92	0.0090
Repres*Param	6.980	1	6.98005	7.05	0.0084
Repres*Lang	0.072	1	0.07184	0.07	0.7878
Param*Lang	0.426	1	0.42588	0.43	0.5124
Error	258.245	261	0.98945		
Total	275.478	267			

Table 28: The results of the ANOVA on Flight A indicate that there are no main effects or interaction effects that impact the number of changes that pilots recommended as a result of their debrief.

Source	**Sum Sq.**	**d.f.**	**Mean Sq.**	**F**	**Prob>F**
Repres	1.574	1	1.57392	0.86	0.3544
Param	5.041	1	5.04114	2.76	0.098
Lang	0.724	1	0.72436	0.4	0.5296
Repres*Param	3.306	1	3.30556	1.81	0.1799
Repres*Lang	2.083	1	2.09344	1.14	0.2867
Param*Lang	0	1	0.0005	0	0.9869
Error	477.216	261	1.82841		
Total	490.985	267			

Table 29: The results of the ANOVA on the risk perception responses from Flight B differ from Flight A's results. We observe that the *Representation Method* and *Parameter Type* factors moved the location of the mean response, but the *Framing Language* factor did not. Rows shaded in the darker gray correspond to parameters that are significant at the 0.05 significance level, with the lighter gray color used to identify rows that came close to the 0.05 significance level.

Source	**Sum Sq.**	**d.f.**	**Mean Sq.**	**F**	**Prob>F**
Repres	8.456	1	8.4559	6.92	0.0092
Param	28.121	1	28.1212	23.01	0.0000
Lang	0.08	1	0.0805	0.07	0.7978
Repres*Param	4.719	1	4.7194	3.86	0.0509
Repres*Lang	0.047	1	0.0472	0.04	0.8445
Param*Lang	0.045	1	0.0452	0.04	0.8476
Error	229.772	188	1.2222		
Total	270.995	194			

Table 30: The results of the ANOVA on Flight B indicate that *representation type* and *parameter type* impacted the number of changes that pilots recommended as a result of their debrief.

Source	**Sum Sq.**	**d.f.**	**Mean Sq.**	**F**	**Prob>F**
Repres	6.016	1	6.0157	7.86	0.0056
Param	8	1	8.00033	10.45	0.0014
Lang	0.167	1	0.1666	0.22	0.6414
Repres*Param	1.363	1	1.363	1.78	0.1836
Repres*Lang	0.243	1	0.24281	0.32	0.5739
Param*Lang	0.866	1	0.86562	1.13	0.2889
Error	143.882	188	0.76533		
Total	160.595	194			

Table 31: As expected, the ANOVA on the risk perception responses from Flight C did not identify any parameters that influenced the results.

Source	**Sum Sq.**	**d.f.**	**Mean Sq.**	**F**	**Prob>F**
Repres	0.012	1	0.01190	0.01	0.9123
Param	0.150	1	0.14959	0.15	0.6961
Lang	0.584	1	0.58448	0.60	0.4404
Repres*Param	1.399	1	1.39905	1.43	0.2331
Repres*Lang	1.587	1	1.58670	1.62	0.2043
Param*Lang	2.379	1	2.37943	2.43	0.1205
Error	177.911	182	0.97753		
Total	184.106	188			

Table 32: The results of the ANOVA on Flight C indicate that there are no main or interaction effects that impact the number of changes pilots say they would make to an upcoming flight.

Source	**Sum Sq.**	**d.f.**	**Mean Sq.**	**F**	**Prob>F**
Repres	2.503	1	2.50293	2.00	0.1593
Param	2.993	1	2.99313	2.39	0.1240
Lang	0.012	1	0.01215	0.01	0.9217
Repres*Param	3.130	1	3.13021	2.50	0.1158
Repres*Lang	0.274	1	0.27442	0.22	0.6404
Param*Lang	0.120	1	0.12006	0.10	0.7573
Error	228.142	182	1.25353		
Total	236.550	188			

Table 33: The distributions of the character length in the total changes suggested are very dispersed, with no clear or consistent differences between them for different factors.

	Changes Character Length							
	Representation Method							
	Graphical				Numerical			
Flight	**Mean**	**Standard Deviation**	**Median**	**IQR**	**Mean**	**Standard Deviation**	**Median**	**IQR**
A	49.6992	48.1149	38	79.75	52.4483	49.0616	42	77.25
B	35.9780	33.7563	29	54.75	53.3077	43.0121	45.5	47.5
C	51.1566	44.0732	45	79	54.4906	44.5015	49	48
	Parameter Type							
	Performance				**Safety**			
Flight	**Mean**	**Standard Deviation**	**Median**	**IQR**	**Mean**	**Standard Deviation**	**Median**	**IQR**
A	55.6084	51.0366	48	79.75	46.1280	45.2311	31	71.75
B	37.4752	32.9067	32	55	53.5426	44.8209	47.5	56
C	48.0968	39.0713	43	66	57.8021	48.4328	50.5	62
	Framing Language							
	Safety-centric				**Risk-centric**			
Flight	**Mean**	**Standard Deviation**	**Median**	**IQR**	**Mean**	**Standard Deviation**	**Median**	**IQR**
A	54.8148	47.4199	49	83.75	47.5038	49.5933	31	73.25
B	46.1584	40.2397	38	66.5	44.2128	39.5643	38.5	48
C	51.4021	44.3278	44	75	54.7391	44.2989	49.5	53.5

Risk Perception Linear Regression

I ran the full complement of all possible factor combinations to estimate all of the main effects between the factors and the results, as well as any interaction effects between factors. The full factorial design will have three main effects, three two-factor interactions, and one three-factor interaction. To model the response variable, I use a linear regression model of the form of Equation 5:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 + \beta_{123} X_1 X_2 X_3 + \varepsilon \qquad (5)$$

The full factorial design will allow us to estimate all eight β_i coefficients $\{\beta_0, \ldots, \beta_{123}\}$. The terms $X_1 X_2$, $X_1 X_3$, and $X_2 X_3$ represent the possible two-order interactions between variables and $X_1 X_2 X_3$ represents the three-order interaction.

The next step in the survey analysis was to therefore fit a linear model to the data and determine whether any of the coefficients indicate main or interaction effects. I fitted three different models; one for each flight, as described by Equation 6. The stepwise linear model started by including all coefficients before removing any coefficients that were not significant to the 95% level one by one. All X variables are categorical and take values of 1 or 0.

$$\begin{aligned} Y = \beta_0 &+ \beta_{\mathrm{Rep}} X_{\mathrm{Rep}} + \beta_{\mathrm{Param}} X_{\mathrm{Param}} + \beta_{\mathrm{Lang}} X_{\mathrm{Lang}} + \beta_{\mathrm{Rep:Param}} X_{\mathrm{Rep}} X_{\mathrm{Param}} \\ &+ \beta_{\mathrm{Rep:Lang}} X_{\mathrm{Rep}} X_{\mathrm{Lang}} + \beta_{\mathrm{Param:Lang}} X_{\mathrm{Param}} X_{\mathrm{Lang}} + \varepsilon \end{aligned} \qquad (6)$$

Table 34: The linear models did not fit the data very well, but they did indicate which factors affect the responses. As expected, the linear model for Flight C only kept the intercept, making it a constant model. The linear model for Flight A found all three factors to be relevant as well as an interaction effect (F-statistic vs. constant model: 4.25, p-value = 0.00238) and Flight B found two of the three factors to have effects, and both of them together produced an interaction effect (F-statistic vs. constant model: 11.4, p-value = 6.79e-07).

Flight	**R^2; R^2(adj)**	**RMSE**		**β_i**	**p-value**	**95% Confidence Interval**
A	0.0607; 0.0464	0.992	β_0	2.63820	1.3125e-53	[2.3742; 2.9022]
			β_{Rep}	0.54795	0.0023542	[0.1967; 0.8992]
			β_{Param}	0.38600	0.0203610	[0.0603; 0.7117]
			β_{Lang}	0.32130	0.0089993	[0.0809; 0.5617]
			$\beta_{Rep:Param}$	-0.65162	0.0079925	[-1.1317; -0.1716]
			$\beta_{Rep:Lang}$	-	-	-
			$\beta_{Param:Lang}$	-	-	-
B	0.152; 0.138	1.1	β_0	3.43140	3.717e-55	[3.1283; 3.7344]
			β_{Rep}	-0.10579	0.641950	[-0.5539; 0.3423]
			β_{Param}	-0.45024	0.037758	[-0.8747; -0.0257]
			β_{Lang}	-	-	-
			$\beta_{Rep:Param}$	-0.62534	0.048743	[-1.2472; -0.0035]
			$\beta_{Rep:Lang}$	-	-	-
			$\beta_{Param:Lang}$	-	-	-
C		0.99	β_0	2.93120	3.3654e-95	[2.7892; 3.0732]
			β_{Rep}	-	-	-
			β_{Param}	-	-	-
			β_{Lang}	-	-	-
			$\beta_{Rep:Param}$	-	-	-
			$\beta_{Rep:Lang}$	-	-	-
			$\beta_{Param:Lang}$	-	-	-
Grouped Flights	0.0716; 0.0571	1.03	β_0	2.7815	1.3819e-84	[2.5417; 3.0213]
			β_{Rep}	0.38193	0.013926	[0.0807; 0.6831]
			β_{Param}	0.26702	0.069301	[-0.0212; 0.5552]
			β_{Lang}	0.17049	0.035927	[0.112; 0.3298]
			β_{F1}	0.63261	0.00015847	[0.3057; 0.9595]
			β_{F2}	-0.039223	0.81244	[-0.3637; 0.2852]
			$\beta_{Rep:Param}$	-0.3577	0.027973	[-0.6766; -0.0388]
			$\beta_{Rep:Lang}$	-	-	-
			$\beta_{Rep:F1}$	-0.64021	0.00108	[-1.0230; -0.2574]
			$\beta_{Rep:F2}$	-0.21001	0.28714	[-0.5971; 0.1771]
			$\beta_{Param:Lang}$	-	-	-
			$\beta_{Param:F1}$	-0.84414	1.6538e-05	[-1.2261; -0.4622]
			$\beta_{Param:F2}$	-0.056523	0.7735	[-0.4420; 0.3290]

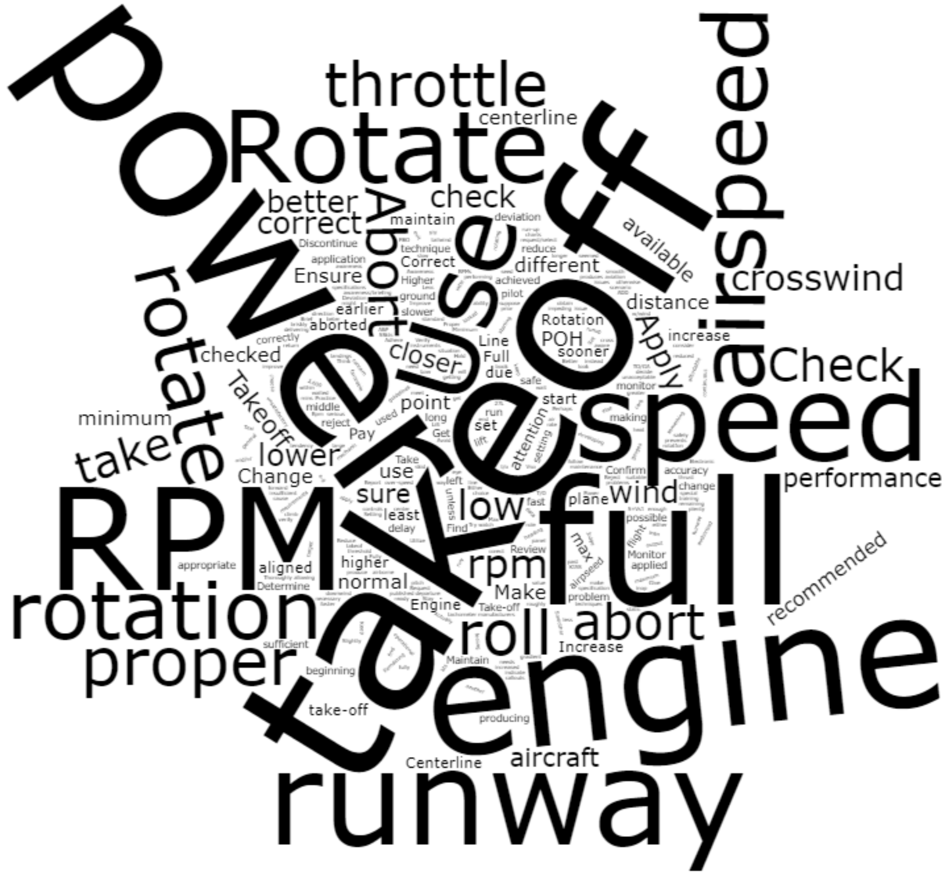


Figure 36: The respondents to Flight B reported that they would make a total of 237 changes to an upcoming flight after reviewing their debrief. The pilots referred to proper takeoff speed and power, and suggested they would take off with full throttle at a higher RPM, or even abort the flight.

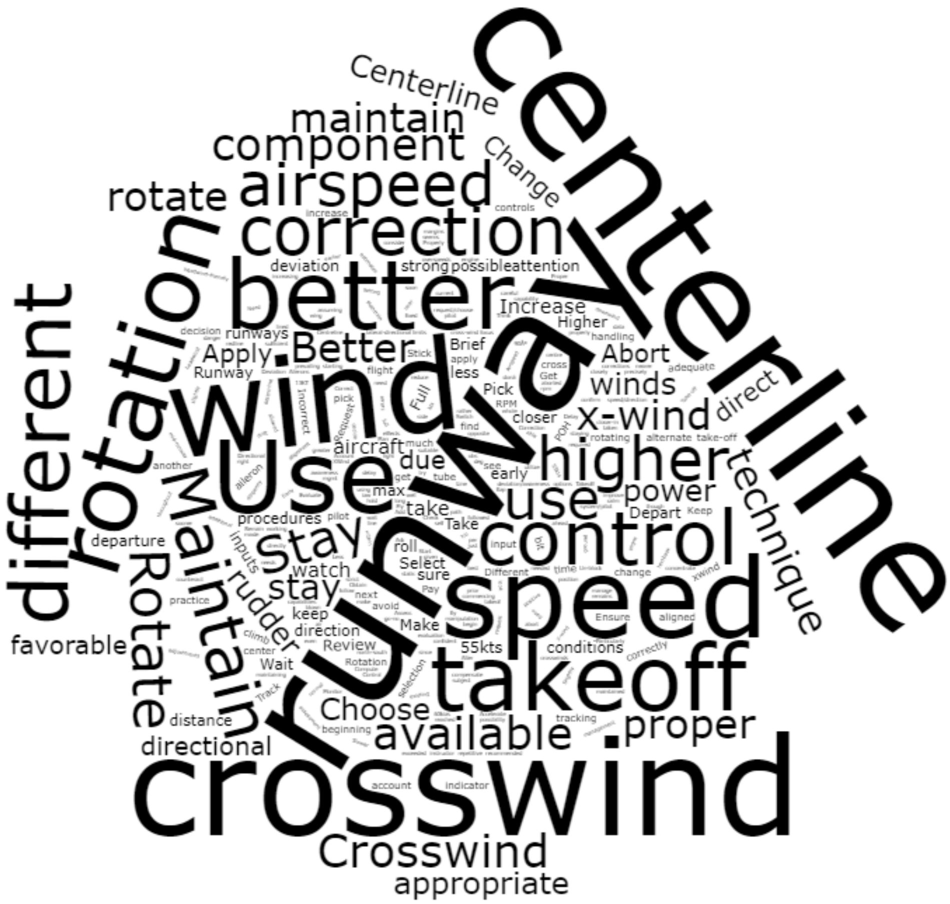


Figure 37: The respondents to Flight C reported that they would make a total of 266 changes to an upcoming flight after reviewing their debrief. The pilots talked a lot about the runway centerline and the crosswind conditions. They also mentioned that the winds favored a different runway and that the appropriate wind correction could keep the plane on the centerline.

REFERENCES

Alaziz, M., Stolfi, A. & Olson, D. M., 2017. Cirrus Airframe Parachute System and Odds of a Fatal Accident in Cirrus Aircraft Crashes. *Aerospace Medicine and Human Performance,* 88(6), pp. 556-564.

AOPA Air Safety Institute, 2018. *2016-2017 GA Accident Scorecard*, Frederick, MD: AOPA.

AOPA Air Safety Institute, 2018. *27th Joseph T. Nall Report General Aviation Accidents in 2015,* Frederick, MD: AOPA.

Babb, T. A., 2017. Electronic Flight Bag Policies at Collegiate Aviation Programs. *International Journal of Aviation, Aeronautics, and Aerospace,* 4(4).

Boeing, 2018. *Pilot Outlook: 2018-2037.* [Online]
Available at: https://www.boeing.com/commercial/market/pilot-technician-outlook/2018-pilot-outlook/
[Accessed 2 March 2019].

Budalakoti, S., Srivastana, A. N. & Akella, R., 2006. *Discovering atypical flights in sequences of discrete flight parameters.*. Big Sky, MT, IEEE Aerospace Conference.

Chakraborty, A., Fala, N. & Marais, K., 2019. *Investigation of Feasibility of Using Low-Cost AHRS Devices to Detect General Aviation Hazardous States.* Dallas, TX, AIAA Aviation Technology, Integration, and Operations Conference.

Chidester, T. R., 2003. *Understanding Normal and Atypical Operations Through Analysis of Flight Data.* Dayton, OH, 12th International Symposium on Aviation Psychology.

Dillingham, G. L., 2013. *Aviation Safety: FAA Efforts Have Improved Safety, but Challenges Remain in Key Areas,* United States Government Accountability Office.

Fala, N. & Marais, K., 2016. *Detecting Safety Events During Approach in General Aviation Operations.* Washington, D.C., AIAA Aviation Technology, Integration, and Operations Conference.

Fala, N. & Marais, K., 2019, *Assessing Potential Biases in Risk Perception for General Aviation Pilots.* Dallas, TX, AIAA Aviation 2019 Forum.

Fala, N. & Marais, K., 2019, *Communicating Data-Driven Risk Information to Pilots.* Dayton, OH, 20th International Symposium on Aviation Psychology.

Fanning, R. M. & Gaba, D. M., 2007. The Role of Debriefing in Smulation-Based Learning. *Society for Simulation in Healthcare*, 2(2), pp. 115-125.

Federal Aviation Administration, 2008. *Aviation Instructor's Handbook.* Washington, D.C.: U.S. Department of Transportation.

Federal Aviation Administration, 2015. *FAA Airport Diagrams.* [Online] Available at: https://www.faa.gov/airports/runway_safety/diagrams/ [Accessed 20 March 2019].

Federal Aviation Administration, 2016. *Pilot's Handbook of Aeronautical Knowledge.* Washington, D.C.: U.S. Department of Transportation.

Federal Aviation Administration, 2017. *Private Pilot - Airplane Airman Certification Standards,* Washington, D.C.: U.S. Department of Transportation.

Federal Aviation Administration, 2018. *FAA Aerospace Forecast Fiscal Years 2018-2038.*

Federal Aviation Authority, 1991. *Aeronautical Decision Making, AC 60-22,* Washington, D.C.: U.S. Department of Transportation.

General Aviation Manufacturers Association, 2006. *GAMA Annual Industry Review & 2007 Market Outlook,* Washington, D.C.: GAMA.

General Aviation Manufacturers Association, 2015. *2015 General Aviation Statistical Databook and 2016 Industry Outlook*, Washington, D.C.: GAMA.

Goblet, V., Fala, N. & Marais, K., 2015. *Identifying Phases of Flight in General Aviation Operations.* Dallas, TX, AIAA Aviation Technology, Integration, and Operations Conference.

Gurm, H. S. & Litaker, D. G., 2000. Framing Procedural Risks to Patients: Is 99% Safe the Same as a Risk of 1 in 100?. *Academic Medicine,* 75(8), pp. 840-842.

Hunter, D. R., 2002. *Risk Perception and Risk Tolerance in Aircraft Pilots,* Washington, D.C. : U.S. Department of Transportation.

Hunter, D. R., 2003. Measuring General Aviation Pilot Judgment Using a Situational Judgment Technique. *The International Journal of Aviation Psychology,* 13(4), pp. 373-386.

International Air Transport Association, 2017. *International Air Transport Association.* [Online] Available at: https://www.iata.org/pressroom/pr/Pages/2017-10-24-01.aspx [Accessed 2 March 2019].

Jamieson, S., 2014. Likert Scales: How to (Ab)use Them. *Medical Education,* Volume 38, pp. 1217-1218.

Jensen, R. S., 1995. *Pilot Judgment and Crew Resource Management.* 1st ed. New York, NY: Routledge.

Jensen, R. S., Adrion, J. & Lawton, R. S., 1987. *Aeronautical Decision Making for Instrument Pilots,* Arlington, VA: Systems Control Technology Inc. .

Jensen, R. S. & Benel, R. A., 1977. *Judgment Evaluation and Instruction in Civil Pilot Training,* Washington. D.C. : Federal Aviation Administration.

Kahneman, D. & Tversky, A., 1973. On the Psychology of Prediction. *Psychological Review,* 80(4), pp. 237-251.

Leong, F. T. & Austin, J. T., 2006. *The Psychology Research Handbook: A Guide for Graduate Students and Research Assistants.* Sage.

Li, L., Gariel, M., Hansman, R. J. & Palacios, R., 2011. *Anomaly Detection in Onboard-Recorded Flight Data Using Cluster Analysis.* Seattle, WA, IEEE/AIAA Digital Avionics Systems Conference.

Lipkus, I. M., 2007. Numeric, Verbal, and Visual Formats of Conveying Health Risks: Suggested Best Practices and Future Recommendations. *Medical Decision Making,* 27(5), pp. 696-713.

Mindstar Aviation, 2016. *Mindstar Avidyne Simulation - Screenshots of the Avidyne Entegra.* [Online]
Available at: http://www.mindstaraviation.com/Avidyne_More_Pictures.html
[Accessed 10 March 2019].

Mitchell, K., Sholy, B. & Stolzer, A. J., 2007. General Aviation Aircraft Flight Operations Quality Assurance: Overcoming the Obstacles. *IEEE Aerospace and Electronic Systems Magazine,* 22(6), pp. 9-15.

MITRE, 2014. *GAARD - General Aviation Airborne Recording Device.* [Online]
Available at: https://www.mitre.org/research/technology-transfer/technology-licensing/gaard-general-aviation-airborne-recording-device
[Accessed 4 July 2017].

National Research Council, 1989. *Improving Risk Communication,* Washington, D.C.: The National Academies Press.

National Transportation Safety Board, 2016. *Most Wanted List of Transportation Safety Improvements,* Washington, D.C.: NTSB.

NextGA Aircraft, Inc., 2012. *Garmin G1000 Details.* [Online] Available at: http://www.nexga.com/columbia-resale-overview/g1000/ [Accessed 10 March 2019].

NGAFID, 2017. *National General Aviation Flight Information Database.* [Online] Available at: https://ngafid.org/ [Accessed 27 December 2017].

O'Hare, D., 1990. Pilots' Perception of Risks and Hazards in General Aviation. *Aviation Space and Environmental Medicine*, 61(7), pp. 559-603.

O'Hare, D., 1992. The 'Artful' Decision Maker: A Framework Model for Aeronautical Decision Making. *The International Journal of Aviation Psychology*, 2(3), pp. 175-191.

O'Hare, D., 2003. Aeronautical Decision Making. In: P. S. Tsang & M. A. Vidulich, eds. *Principles and Practice of Aviation Psychology.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc., pp. 201-237.

Paling, J., 2003. Strategies to Help Patients Understand Risks. *BMJ*, 327(7417), pp. 745-748.

Rao, A. H., 2016. *A New Approach to Modeling Aviation Accidents.* West Lafayette: Purdue University.

Ray, B., 2013. *High Performance Aviation, LLC - Garmin G1000 VNAV Tips and Tricks.* [Online] Available at: https://www.flyhpa.com/2013/01/garmin-g1000-vnav-tips-and-tricks/ [Accessed 10 March 2019].

Salas, E. et al., 2008. Debriefing Medical Teams: 12 Evidence-Based Best Practices and Tips. *The Joint Commission Journal on Quality and Patient Safety*, 34(9), pp. 518-527.

Shappell, S. A. & Wiegmann, D. A., 1997. A Human Error Approach to Accident Investigation: The Taxonomy of Unsafe Operations. *The International Journal of Aviation Psycholoy,* 4(4), pp. 269-291.

Sullivan, G. M. & Artino, A. R. J., 2013. Analyzing and interpreting data from Likert-type scales. *Journal of Graduate Medical Education,* 5(4), pp. 541-542.

Sumwalt, R. L. & Dalton, S. L., 2014. *The NTSB's Role in Aviation Safety*, Orlando, FL: National Business Aviation Association Emergency Response Planning Workshop.

Tannenbaum, S. I. & Cerasoli, C. P., 2013. Do Team and Individual Debriefs Enhance Performance? A Meta-Analysis. *Human Factors,* 55(1), p. 2310245.

Tversky, A. & Kahneman, D., 1974. Judgment Under Uncertainty: Heuristics and Biases. *Science,* 185(4157), pp. 1124-1131.

United States Government Accountability Office, 2018. *Collegiate Aviation Schools: Stakeholders' Views on Challenges for Initial Pilot Training Programs*.

Utah State University, 2017. *Utah State Universtity Aviation Technology 2018-2019 Student Handbook.* [Online]
Available at: https://aste.usu.edu/aviation/files/Student%20Handbook%202018_2019.pdf
[Accessed 3 March 2019].

Veillette, P. R., 2014. Team LOSA, ASAP, and FOQA, and How They Supplement Each Other.. *Business & Commercial Aviation*, 1 June.

Wright, K. B., 2005. Researching Internet-Based Populations: Advantages and Disadvantages of Online Survey Research, Online Questionnaire Authoring Software Packages, and Web Survey Services. *Journal of Computer-Mediated Communication,* 10(3).

VITA

Nicoletta Fala completed all of her degrees and flying licenses/ratings at Purdue University. She received her Bachelor’s and Master’s degrees in Aeronautics and Astronautics at Purdue University in 2014, with a focus on Aerospace Design and Systems respectively.

Nicoletta worked in the VRSS (Value through Reliability, Safety, and Sustainability) Lab at Purdue University under the guidance of Prof. Karen Marais for five years, and gained experience in systems engineering and risk assessment of complex systems in general, with a focus on general aviation safety, human factors in transportation systems, and risk communication.

While at Purdue, Nicoletta was an active member and officer of Purdue Pilots, Inc., serving as the club’s president for a year, and got her Commercial Pilot License, Instrument Rating, and Certified Flight Instructor rating while pursuing her PhD. She participated in the Air Race Classic twice (2017 and 2018), and took every opportunity to fly in all kinds of terrain and weather systems.