AN ANALYSIS OF FEDERAL AVIATION ADMINISTRATION KNOWLEDGE TEST SCORES AND FATAL GENERAL AVIATION ACCIDENTS

by

BERNARD FRANCIS KING

B.S., United States Military Academy, 1969
M. Ed., Georgia State University, 1975
M.A., Webster University, 1998

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Special Education, Counseling and Student Affairs
College of Education

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2015

Abstract

Over the last few years, the safety record of U.S. commercial airlines has improved to the point where the statistics on accidents are negligible. The overwhelming numbers of aviation fatalities occur in General Aviation (GA) accidents. While the fatal accident rate has improved—from around 5.0 per 100,000 miles flown in the post–World War II era to varying between 1.2 and 1.5 since 1996—it still results in 450 to 700 deaths per year. In 2013, improving GA safety was on the National Transportation Safety Board's (NTSB's) most wanted list.

The NTSB has cited a lack of aeronautical knowledge as the cause of many of these accidents. If pilots are required to pass Federal Aviation Administration (FAA) knowledge and practical tests prior to obtaining a new pilot certificate, how could they not possess the knowledge needed to operate in the National Airspace System (NAS)? Some, attributing it to a failure to learn basic aeronautical knowledge, are concerned that potential pilots memorize the answers to test questions published in commercially available test guides and quickly forget the material after passing the test.

The purpose of this retrospective causal comparative study was to see if airman knowledge tests scores are related to fatal accidents. Fatal GA accidents that had pilot error as a causal factor were compared with those in which maintenance factors caused the accidents, to see if there were significant differences in the pilots' knowledge test scores. The time that potential pilots took to answer skill-based questions that required calculation or interpretation was compared to the time to answer recall questions to see if rote memorization may have been involved in passing the knowledge test.

The results of this paper may have implications on how AFS 630 structures the FAA knowledge tests and how instructors prepare potential pilots for these tests.

AN ANALYSIS OF FEDERAL AVIATION ADMINISTRATION KNOWLEDGE TEST SCORES AND FATAL GENERAL AVIATION ACCIDENTS

by

BERNARD FRANCIS KING

B.S., United States Military Academy, 1969
M. Ed., Georgia State University, 1975
M.A., Webster University, 1998

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Special Education, Counseling and Student Affairs
College of Education

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2015

Approved by:

Major Professor
Fred O. Bradley

Abstract

Over the last few years, the safety record of U.S. commercial airlines has improved to the point where the statistics on accidents are negligible. The overwhelming numbers of aviation fatalities occur in General Aviation (GA) accidents. While the fatal accident rate has improved—from around 5.0 per 100,000 miles flown in the post–World War II era to varying between 1.2 and 1.5 since 1996—it still results in 450 to 700 deaths per year. In 2013, improving GA safety was on the National Transportation Safety Board's (NTSB's) most wanted list.

The NTSB has cited a lack of aeronautical knowledge as the cause of many of these accidents. If pilots are required to pass Federal Aviation Administration (FAA) knowledge and practical tests prior to obtaining a new pilot certificate, how could they not possess the knowledge needed to operate in the National Airspace System (NAS)? Some, attributing it to a failure to learn basic aeronautical knowledge, are concerned that potential pilots memorize the answers to test questions published in commercially available test guides and quickly forget the material after passing the test.

The purpose of this retrospective causal comparative study was to see if airman knowledge tests scores are related to fatal accidents. Fatal GA accidents that had pilot error as a causal factor were compared with those in which maintenance factors caused the accidents, to see if there were significant differences in the pilots' knowledge test scores. The time that potential pilots took to answer skill-based questions that required calculation or interpretation was compared to the time to answer recall questions to see if rote memorization may have been involved in passing the knowledge test.

The results of this paper may have implications on how AFS 630 structures the FAA knowledge tests and how instructors prepare potential pilots for these tests.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I thank those who have contributed to this dissertation over the last few years. I was looking for a topic that would make a contribution to my day job in the aviation department at KSU but also be education-related, so as to satisfy my degree requirement in higher-education student affairs.

On a trip to the Mike Monroney Aeronautical Center with my altitude chamber class, I scheduled a meeting with Dennis Byrne, the Educational Program Specialist in AFS 630. He provided a briefing on AFS 630 and FAA knowledge testing. He provided the basis for this study when he indicated the need for a study on FAA knowledge testing and safety. He and the staff were also instrumental in gathering accident data from the NTSB and matching knowledge-test results for the pilots involved in these accidents.

I thank my committee members, starting with my major professor Fred Bradley, R. Kurt Barnhart, Sheryl Hodge, and Christy Craft, who have provided continued guidance, support, and patience as I progressed through this process. Finally, I thank Doris Carroll, with whom I served on the Academic Affairs committee, for getting me started on this degree. I was working on a graduate certificate in advising when Doris took the time to show me the options in applying that certificate toward a degree; she has encouraged me along the way.

Dedication

This dissertation is dedicated to my wife of 43 years, Mary Ellen, whom I have often neglected as I have pursued military diplomas and advanced degrees. Thank you for allowing me the time to follow my academic aspirations.

Chapter 1 - Introduction

Improving GA safety is one of the items on the National Transportation Safety Board's (NTSB's) most wanted list. While the major airlines have few fatal accidents, the NTSB continues to investigate approximately 1,500 GA accidents each year. Among many other contributing/causal factors in these mishaps, the NTSB cites a lack of knowledge, skills, and recurrent training. It notes that:

> Knowledge tests and flight reviews should test for awareness of weather, use of instruments and use of "glass" cockpits. And there should be a mechanism for identifying at-risk pilots and addressing risks so that both pilots and passengers can safely fly. (p.2) (National Transportation Safety Board, 2012)

According to Campbell and Bagshaw (2002), more that 70% of GA accidents can be attributed to human error. They pointed out that pilots must have technical knowledge in meteorology, aircraft engines, instruments, and navigation, and also must exercise judgment, which is based on knowledge, training, and experience.

Many studies have been done concerning knowledge as it relates to student learning. Farr (1987) indicated that the greater a student learns material or over-learns it by having additional practice with it, the slower will be the rate of decay. Students should also learn concepts, principles, and rules to supplement rote learning, enabling them to recall details that they would not otherwise remember. Other authors have studied the amount of knowledge a person learns and the rate of decay. Casner, Heraldez, and Jones (2006) found that material memorized by rote is weaker than material processed in a more meaningful way. Ellis, Semb, and Cole (1998) also found that material must be well learned to be remembered over time. Knecht (2010) acknowledged another issue that may affect aviators: what is learned in the classroom may not transfer to the real world of flight. He indicated that studies conducted in classrooms may not be transferable to the flight line. Some feel that the problem lies with the transfer of knowledge from the educator to the pilot. Stowell (2008) did not feel that aviators could be held accountable for a lack of knowledge leading to an operational error if they never learned the material.

Accidents can be broken down roughly into human and mechanical failures. Airplanes are becoming increasingly more reliable and technically advanced. Amalberti and Wibaux (2000) pointed out that accident rates are not decreasing, even though there are fewer dramatic mechanical failures—but the rate of pilot cognitive errors is increasing.

This issue is not unique to the United States. Lenne, Ashby, and Fitzharris (2008) studied Australian aviation accidents and found similar pilot error rates. They also found that skill-based and decision errors were more likely if the pilot had personal readiness issues of an adverse physical or mental state.

In their studies, Casner et al. (2006) found that regular flying or teaching and biannual flight reviews do not keep pilots familiarized with all they have learned. They suggested the establishment of some type of standards for ongoing aeronautical knowledge proficiency.

The Federal Aviation Administration (FAA) AFS 630, is responsible for determining if airmen have the knowledge and skills to operate in the National Airspace System (NAS). The agency develops all airman knowledge test guides, knowledge test questions, and practical test standards (PTS) (Federal Aviation Administration, 2012a).

The FAA publishes example questions for the variety of knowledge tests used for airman testing. Casner, Jones, Puentes, and Irani (2004) indicated that the FAA is concerned that, if it published all the questions, applicants would concentrate only on the questions asked in the test bank, which might leave gaps in their total aviation knowledge. Because they learned the questions merely by rote, they would not understand the concepts on which the questions are based.

Statement of the Problem

What can be done to keep pilots from killing themselves and their passengers? Studies have shown who is most likely to be involved in an accident. Trollip and Jensen (1991) found that, according to NTSB records, the mostly likely candidate for an accident is 35 to 39 years old, has 100 to 500 hours of flight time, and is on a personal flight in visual meteorological conditions (VMC). The NTSB (2012) has cited lack of knowledge as a primary or contributing cause of many of these accidents. There are several possible causes of this lack of knowledge. Pilots may never have learned the material—but with written knowledge tests and oral and practical examinations, it seem unlikely that they could pass these exams without the minimum basic knowledge required for the rating. If they were successful in attaining their pilot certificates, what could have caused them to forget or not apply the material? Retention of learned knowledge decreases over time. Farr (1987) determined that deep learning slowed the rate of decay. Could it be that many student pilots are passing the FAA knowledge test by rote memorization of questions instead of really understanding the material? This could lead to rapid decay of knowledge and resulting higher accident rates.

In the early days of aviation, flying was considered a rule-based skill. Harris and Muir (2005) pointed out that flying is now more knowledge-based. The increase in technology has caused flying to be more complex, which can sometimes lead to delayed responses and, therefore, errors. As a result of that increased complexity, more training is generally required to achieve a satisfactory level of performance.

Among GA pilots, there is a large difference in accident rates between corporate pilots who fly regularly and receive regular recurrent training and pilots who do not receive regular training. Hopkins (1994) determined that pilots who received regular check rides generally had an accident rate of fewer than one accident per 100,000 miles flown, whereas those who receive no regular training have an accident rate near seven accidents per 100,000 miles. He also noted that attitude has an effect on flying. Pilots with the attitude to barely meet the FAA minimum standards to get their ratings and to never improve or upgrade their skills often end up as statistics.

It has also been determined that accidents involving flight into instrument meteorological conditions result in more fatalities than do an equal number of accidents occurring in visual conditions. W. Knecht, Harris, and Shappell (2005) found that approximately 32% of fatalities involved instrument meteorological conditions.

Purpose of the Study

How can the loss of lives in GA accidents be prevented? It is known that many accidents are caused by lack of knowledge. Is it possible that pilots are able to gain a pilot certificate without the required knowledge, or does their knowledge base decay over time due to normal forgetting? Does depth of learning acquired by superficial or rote learning for the test contribute to a faster rate of knowledge decay?

Questions on the FAA knowledge tests generally fall into two categories. The first is define or recall, which can be answered by rote learning. The other type requires calculation or interpretation. If a student were able to pass knowledge test by memorizing answers to calculation problems without understanding the material, would he or she be a safe pilot? If the student encountered a situation that was slightly different than the scenario in the learning environment, could that student apply the knowledge that he or she possesses?

The FAA has already realized that the knowledge and PTS need to be aligned with each other (Federal Aviation Administration, 2012b). The agency has tasked the current Aviation Rulemaking Advisory Committee (ARAC) with ensuring that the knowledge questions are consistent with the aligned test standards and the principles developed by the previous Aviation Rulemaking Committee (ARC).

The purpose of this study is to contribute to the body of knowledge concerning airman knowledge testing as it relates to pilot involvement in GA aircraft accidents. It is hoped that this will add to what is known about the pilots who are most likely to be in an accident. If it is possible to identify such pilots, it may be possible to make changes to our testing methods to prevent them from obtaining a pilot certificate without sufficient knowledge or sufficient depth of knowledge.

Background

GA pilots are involved in more accidents than are airline pilots. Hackworth et al. (2007) gave a number of possible reasons for this. GA pilots generally make more takeoffs and landings, which are higher risk maneuvers. Other high-risk operations include aerial application, flight instruction, and flights from less-improved runways. GA pilots are also, on the whole, less experienced than airline pilots. Campbell and Bagshaw (2002) noted that aviation training utilizes knowledge, training, and skill. Knowledge has to be gained before skill can be developed and subsequently improved with experience. Croft (2012) is of the opinion that pilots often do not have the knowledge, skills, or recurrent training to fly safely, particularly in questionable weather conditions. When pilots are in the air, they have limitations which can contribute to them committing errors. Rolfe (1972) pointed out that human factors research shows that man's senses are restricted; he has limited ability to process information, and there are delays before he is capable of responding. In studies concerning locus of control, it was found that pilots are more internal—that is, they believe they can influence outcomes by their own efforts—than people in the general population, who believe outcomes are more due to chance (Stewart, 2008). Pilots were also found to be biased toward their own skills and safety. They thought of themselves as more skilled and safer than pilots with similar numbers of flight hours. The author suggested that internality and self-confidence can promote behavior that promotes safer flying.

Year after year, the NTSB finds accidents with similar recurring circumstances. The FAA has acknowledged this problem for a long period of time. Cole (1967) found that most GA accidents were caused by ignorance, complacency, and inadequate flight training. The FAA tried, unsuccessfully, to mitigate these problems in 1965 with the General Aviation Pilot Education (GAPE) program, which was based on a statistical analysis of accidents occurring in 1964 and 1965. Information about then-current trends in mishaps was brought to the attention of the GA community. Like many of the other programs attempted by the FAA through the years, the result was no appreciable change in the fatality rate or numbers of accidents.

The current FAA knowledge and proficiency program designed to address causal factors associated with common pilot errors is the WINGS program (Federal Aviation Administration, 2011a). Development of this Internet-based knowledge and skill program was based on a study of two years of accident data in which the causal factors were compiled by category and class. Tasks designed to address these causal factors are designed either as knowledge tasks or flight activities. The program has three levels: basic, advanced, and master. Knowledge tasks can be accomplished online or at safety meetings sanctioned by the Federal Aviation Administration

Safety Team (FAASTeam). The flight activities must be done with a certified flight instructor in accordance with the PTS. The flight instructor must go online and verify the activity before the airman receives credit for the activity. A pilot must complete all the tasks for a phase of the WINGS program within 12 months. Once a phase of the program is completed, it is valid for 12 months.

The FAA assembled a panel of experts in 2012 and formed an Airman Testing Standards and Training ARC with the following charge: (Hennig, 2012)

The FAA asked the ARC to review and provide recommendations in four areas:

1. *A prioritized list of up to five pilot and/or instructor certificates and/or ratings its work will address.*
2. *An aeronautical knowledge standard for the selected certificates and ratings. The aeronautical knowledge standard for each certificate and/or rating should set forth the overall precepts that will conceptually frame, guide, and justify its specific technical subject areas.*
3. *Methods for regular industry participation in the planning, development, production, and review of technical information (e.g., training handbooks, knowledge test guides, and supplements) intended to convey the elements of the knowledge standard.*
4. *Precepts for development and appropriate review of updated knowledge tests that will accurately and reliably measure the airman's mastery of the aeronautical knowledge standard. This task should include recommendations on types of questions to be included.* (p. vii)

As a result of the ARC, the FAA tasked the ARAC with the following: Combine 14 requirements from CFR Part 61 for knowledge and practical tests for the private, certified flight-instructor and instrument ratings into a single document for each type of rating, realign the FAA handbooks with the new standards, and propose a new set of knowledge test questions that meet the criteria developed by the ARC.

Research Questions

In relations to the causes of aviation accidents, this research is guided by the question: how can the number of GA fatal accidents be reduced?

1. Do pilots involved in fatal accidents (due to human error) have lower scores on FAA knowledge tests than pilots involved in fatal accidents due to mechanical failures or with undetermined causes?
2. Do pilots involved in fatal accidents spend less time per question on FAA knowledge tests than pilots involved in fatal accidents due to mechanical failures or with undetermined causes?
3. Do pilots involved in fatal accidents due to human factors errors spend less time on recall questions than pilots involved in maintenance factors accidents.
4. Do pilots involved in fatal accidents due to maintenance factor spend more time on calculate questions than pilots involved in human factor error accidents.

Hypothesis

Pilots involved in fatal GA aircraft accidents involving pilot error are more likely to have been involved in rote learning of FAA questions than pilots dying in maintenance-related accidents.

Scope and Limitations of the Study

This study was limited in the following ways:

1. This study used data from those pilots involved in fatal accidents. It is possible that they do not accurately represent the total population of those involved in accidents or the general pilot population. Maurino (1999) pointed out that, in using data from accidents, failed behavior data are being used when studying the balance between mission accomplishment and safety. He pointed out that it might be better to gain insights into pilot behavior by placing observers in aircraft and looking at flights with positives outcomes for the keys to avoiding incidents or accidents. This has been done in the airlines by placing trained observers in the cockpit with checklists and rating the observable actions that have been implicated in accidents or incidents. This error management has been documented by Helmreich and Merritt (1998). The age of the pilots was compared to other FAA data to see if the ages of pilots who had accidents were representative of the average age of pilots in the GA population.
2. It is also known that the proportion of accidents associated with violations is considerably higher for fatal accidents than for non-fatal ones. The NTSB summaries were read to see if any of the accidents appeared to be the results of willful violations.
3. In trying to link multiple-choice questions to safety, Hubbard (2000) believed that these questions cannot provide an adequate link to psychomotor skills. Previous studies have shown that motor skills account for a small portion of accidents, with cognitive and perceptual errors being far more likely causes.

4. Another difficult-to-quantify component that can lead to accidents is judgment. In his studies, Hunter (2003) looked at a situational judgment test (SJT) as a means of determining pilot judgment outside of an actual flight in an aircraft or simulator. He believed that it might be useful in certification testing of pilots. This is supposed to be covered during practical tests, but it is difficult to do because of cost and safety. You may not want to put a pilot in a situation (such as an engine failure at low altitude) to test his judgment when a miscalculation could put the examiner in a position where he could not recover without a mishap. Driskill, Weissmuller, Quebe, Hand, and Hunter (1998) pointed out that, whatever testing method may be developed to measure judgment certification process, there was no way to assure that pilots would react and make similar judgments when faced with the stress and motivations of actual flight conditions.
5. A weakness of the study was a lack of randomization. The two groups were naturally formed. It may be possible to create equal groups by matching their flight ratings. A private pilot from one group could be matched with a private pilot from the other group. It may also be possible to create subgroups by flight ratings. Because previous studies have shown that about four times more accidents were due to human error than mechanical factors, this might result in many data not being used. These groups are similar to using intact classroom statistics, in which the means have been used successfully.

Significance of the Study

Craig (2013) pointed out that it is possible to predict, with reasonable accuracy, the number of GA pilots who would be killed in a year and what they would be doing when they died. He posed the question of whether or not it is possible to teach a pilot out of the next accident. Up to the present time, there has not been any linkage between knowledge test scores and an increase in propensity for low scorers to be involved in a greater number of accidents (over those who scored high on the test) (Hennig, 2012). Other studies have been conducted on gender, with Vail and Ekman (1986) coming to the conclusion that women are involved in fewer accidents and suffer fewer fatalities in the accidents in which they are involved. When it comes to preparing for knowledge test, Trollip and Jensen (1991) made these four observations: Studying for the test will probably result in your remembering the information for only a short period of time. Rote memorization will not ensure that you understand the material. If circumstances are different, you will not have the skills or information to deal with it. Finally, tests do not cover all the information that needs to be known, and you may miss an important point.

Assumptions

1. The sample of pilots involved in fatal accidents is representative of the GA population.
2. The causes of some of the accidents investigated by the NTSB could not be determined and were not used in this study. It is assumed that the accidents not included in the study would not change the results.
3. Knowledge test results are a normal distribution, on which parametric tests will be able to be run.

Definitions of Terms

AFS 630 is the Airman Testing Standards Branch of the Regulatory Support

Division of the FAA located in Oklahoma City, OK.

Check rides is a term used by pilots for a variety of recurrent training evaluations or practical tests leading to additional pilot ratings.

The Human Factors Analysis and Classification System is a taxonomy that describes the human factors that contribute to an accident or incident.

Instrument Meteorological conditions are those of reduced visibility and cloud height above the ground that require additional flight training beyond that required for visual flight.

The Situational Judgment Test is a test that exposes the taker to a series of hypothetical situations and asks the taker to pick the most appropriate response or to rank a series of responses in the order of their effectiveness.

A flight review consists of an hour of ground training and an hour of flight training conducted by an instructor within the previous 24 calendar months.

The WINGS program is an FAA-developed voluntary pilot education and proficiency program.

Level C and D simulators are the most sophisticated levels of simulation used by the airlines and some turbojet and turbo prop corporate aircraft.

A flight-simulation training device does not require a motion system and, for the less sophisticated versions, may have a generic airplane cockpit.

Abbreviations

ACS Airman Certification Standards

ADS-B Automatic dependent surveillance-broadcast

AMT Aviation Maintenance Technician

AOPA Airplane Owners and Pilots Association

ARAC Aviation Rulemaking Advisory Committee

ARC Aviation Rulemaking Committee

ATP Airline Transport Pilot

CASA Australian Civil Aviation Safety Authority

CAT Computerized Adaptive Testing

CATS Computer Assisted Testing Services

CFI Certified Flight Instructor

DPE Designated Pilot Examiner

FAA Federal Aviation Administration

FAASTeam FAA Safety Team

FBO Fixed Base Operator

FIRC Flight Instructor Refresher Course

FITS FAA/Industry Training Standards

FSTD Flight Simulation Training Device

GA General Aviation

GAPE General Aviation Pilot Education

HFACS Human Factors Analysis and Classification System

IACRA Integrated Airman Certification and Rating Application

IEEE Institute of Electrical and Electronics Engineers

IFR Instrument Flight Rules

IMC Instrument Meteorological Conditions

IRA Instrument Rating Airplane

IRB Institutional Review Board

JAA Joint Aviation Authority

LCU Life Change Unit

MAC Mid Air Collision

MANOVA Multivariate Analysis of Variance

NACO National Aeronautical Charting Office

NASA National Aeronautics and Space Administration

NAS National Airspace System

NCSBN National Council of State Boards of Nursing

NTSB National Transportation Safety Board

PTS Practical Test Standards

SATS Small Aircraft Transportation Systems

SJT Situational Judgment Test

SRM Single Pilot Resource Management

TAA Technically Advanced Aircraft

VFR Visual Flight Rules

VMC Visual meteorological conditions

WASP Women Air Force Service Pilots

Chapter 2 - Literature Review

The literature review consists of five parts. The first part deals with learning. Both rote and deep learning and the characteristics of each are examined. In the second part, both declarative and procedural knowledge are discussed. The third part is devoted to error and how this can lead to accidents. In the fourth part, aviation safety—including accident rates, trends, and age and gender comparisons—are examined. The fifth part is devoted to knowledge testing and linkages to safety.

There have been many cross-sectional studies in aviation safety research, but there are few controlled studies, such as case-control or cohort studies. Li (1994) noted that, unlike studies of diseases that have mortality tables, there is no such guide for measuring aviation crashes. Most aviation studies have been based on pilot performance criteria, such as psychomotor skills, memory, judgment, decision making, or having been involved in an aviation accident.

Learning

Learning can be either rote or deep. According to Billing (2007), motivation is the key to which type of learning occurs. He also noted that, if students set self-improvement and learning goals, they are more likely to engage in cognitive activities that will improve learning. The aviation industry often resorts to the use of mnemonics to memorize important procedures. Moore and Telfer (1990) pointed out that this reflects a surface or rote approach to learning.

Rote

Learned material is affected by how it is initially learned (Casner et al., 2006). Information that is memorized in a rote manner cannot be remembered as well as material that has been studied and is more meaningful (Craik and Lockhart, 1972). This led Casner et al.

(2004) to believe that some pilots try to memorize FAA questions to pass the test and that they may not understand the material underlying the question. Moore and Telfer (1990) believed that surface learners were merely trying to meet the minimum requirement; thus, they used rote techniques. In addition, these learners were often able to remember many details but there was little structure to their knowledge. Farr (1987) said that concepts should be taught to supplement or complement rote learning, allowing students to generate details that would not otherwise be available to them. Other studies have shown that rote learning is not conducive to long-term memory and, more importantly, that it can decrease a person's ability to apply what he or she has learned (Moebus et al., 2009). Evidence of rote learning can be seen in the time taken to answer questions on FAA knowledge tests. In some cases, students took less time to answer the questions on a test than it would take the average student to read the question.

There have been some efforts to look at using quizzes as a learning tool instead of just evaluation. Stefansson and Sigurdardottir (2011) looked at systems where students can request additional quizzes until a satisfactory result is attained. While this might appear to be rote learning, some of the questions were repeatedly assigned with random numbers to distinguish between rote memorization and learning with understanding.

The Government Accountability Office (2011) pointed out that there have been concerns from some stakeholders that the current knowledge tests are based on rote memorization. There is a concern that pilots focus on memorization rather than developing a true understanding of the material.

Deep

Studies have shown that, for deep learning to occur, the material must be made available for multiple practice sessions and that it must be presented in multiple ways (Ellis, Semb, and

Cole, 1998). This high level of initial learning allows a student to make better selections many years later. Deep learners tend to personalize their learning and look for activities that give meaning to their study (Moore and Telfer, 1990).

Some authors conclude that it is not the difference between rote and deep learning that is the problem that causes accidents, but the pilot's inability to assess his or her own level of knowledge (Amalberti and Wibaux, 2000). The authors pointed out that the gap tends to disappear with experience but that it generally takes at least 600 hours or more than a year of concentrated airline flying. Craig (2013) came up with a very similar figure among his study of GA pilots. Moore and Telfer (1990) concluded that rote-learning strategies were harmful to overall performance. He encouraged deep learning by allowing students to self-question and monitor their own learning in an anxiety-free environment.

Fata-Hartley (2011) concluded that student-centered learning resulted in increased student learning. For students to retain knowledge, they must be engaged. The author put it very well when he said that learning is not a spectator sport.

Bloom's Taxonomy

Studies of successful pilots have shown that the elements of Bloom's taxonomy that are most often seen are analysis, synthesis, and evaluation (Hubbard, 2000), all of which are upper-order elements. Pilots are planning, preparing, or performing tasks from the initiation of the flight until the final paperwork is completed. Lythgoe (2011) studied the analysis process used by Captain Sullenberger when he struck a flock of geese after taking off from La Guardia airport in 2009. His first instinct was to return to his departure airfield. A quick mental simulation proved this was not feasible, nor was his second choice of Teterboro Airport. His third option was the Hudson River, which provided a large flat area to land the aircraft, but landing in water would require a perfect touchdown just above stall speed with the nose slightly high and a minimal rate of descent. He was able to safely land the aircraft and direct the evacuation of the passengers with no fatalities. As Hubbard pointed out, successful pilots must have a high order of analysis and evaluation skills; Captain Sullenberger certainly demonstrated these characteristics.

Knowledge

De Jong and Ferguson-Hessler (1996) presented an in-depth article on types and depth of knowledge. The authors made a partial list of all the terms that have been used to describe knowledge. It would seem that different authors needed many terms to describe knowledge. In this paper, Konoske and Ellis's (1991) definition is used. They divided knowledge into two parts: declarative and procedure. Declarative knowledge is facts or knowledge about things, while procedural knowledge is how to do something. Adams (1992) indicated that, to demonstrate expert cognitive knowledge, a pilot must have a well-ordered body of conceptual and procedural knowledge and be able to access it rapidly. The expert is able to retrieve the information more readily than novices can; the expert is able to reduce the role of memory search and general processing and, because his or her knowledge is more structured, the expert is better able to retrieve and recognize patterns and inference. This is especially important when receiving sensory signals. Campbell and Bagshaw (2002) pointed out that the signals must be interpreted, and pilots must have knowledge from past study and experience in order to correctly process those signals.

It is generally accepted that takeoff and landing constitute the times when pilots have the heaviest workload. Schvaneveldt, Beringer, Lamonica, Tucker, and Nance (2000) were able to document that it is also the time when the pilot has the requirement for more data per unit of time. In their study, they were concerned with finding the priority data requirements in each phase of flight and determining how experienced and novice pilots compared in their assessment of priorities. Both experienced and novice pilots rated the approach and the landing as the highest priority and most demanding phases of flight.

There have been many traditional laboratory studies on learning and retention. In many of these, there was a rapid loss of knowledge. Semb and Ellis (1994) found that the forgetting curve for knowledge gained in the classroom was not nearly as steep as that found in laboratory studies. They also found that, by increasing the amount of original learning, one can increase retention. Higher-ability students are able to learn more than lower-ability students but, the authors did not find any difference in the rate of forgetting.

Casner et al. (2006) study of retention of knowledge yielded the following observations. Recent flight experience appeared to be of more significance than total flight experience, especially for flight instructors. Having recently been involved in a flight review or in preparation for a practical test seemed to create the climate for better retention. Finally, better retention was found in pilots who reported reading more aviation-related materials.

During World War II, the military conducted many studies on pilot judgment and the causes of accidents. Jensen (1995) reported that five factors were repeated by 23% of the pilots with multiple accidents. These factors were:

- A below-average flight record.
- A below-average adaptability for flight.
- An excessive grounding history.
- Lapses of memory and concentration, resulting in incidents.
- Less education in the accident group.

Jensen went on to state that knowledge is essential for coping with emergencies. Going beyond, he said that understanding, more than knowledge, is required to fly safely. You can fly without understanding, but you cannot fly safely without it.

In her dissertation about accidents and fatalities in World War II, Pierce (2013) found evidence that the Aircrew Classification Battery, which was used to determine who became pilots, navigations, or bombardiers, also was a good indicator of pilot-error accidents. Those with the highest scores were less likely to be involved in this type of accident. Unfortunately, in GA, anyone can train who can afford to learn to fly—not merely those who would be the best candidates.

A significant finding was advanced by Kruger and Dunning (1999), who found that knowledge was important in most domains and that people differed in the amounts of knowledge that they possess. Their third and most controversial point was that many people made mistakes and suffered from their decisions, but they did not realize that they were wrong and continued to think that things were fine. This could result from the task being very complex or the person not having the knowledge to perform the task.

Orasanu, Martin, and Davison (2001) noted that, in many accidents, a lack of relevant knowledge contributed to the accident by leading the crew to make poor decisions. They also believed that lack of current knowledge was an issue. Pilots who recently transitioned to another aircraft do not have a great deal of relevant experience, and those who have been away from flying for a period of time may lack current knowledge. In some situations, this lack of relevant or current knowledge can lead crews to underestimate risks. They may also base decisions on past experiences, where they were able to land in poor weather at a certain airport one time; they might expect that, even if the conditions are not good, they might be able to do it again.

W. R. Knecht and Lenz (2010) called this confidence calibration. Many pilots overestimate their ability and judgment and have more confidence in being able to handle situations than their current performance warrants. They think they can put forth a maximum effort whenever they need it, without looking at factors such as physical health, nutrition, and sleep that may prevent them from operating at 100% at any time.

Another argument advanced by Amalberti (1993) is that it is not the difference between superficial and deep knowledge that effects aviation safety but the pilot's ability to estimate his level of knowledge. He indicated that this is especially true for beginners, which he said lasted until a pilot had accumulated about 600 flight hours.

Woods, Johannesen, Cook, and Sarter (1884) had a similar theory, which they called knowledge calibration. They indicated that people are well calibrated if they are aware of how well they know what they know. The authors said that people are miscalibrated if they are overconfident or lacked certain knowledge or if the knowledge they possessed was buggy. This is their term for incomplete or erroneous knowledge.

Declarative

A common theme that Craig (2013) found in the accidents he reviewed was that you must know the aircraft you fly in order to fly it safely. This should be not just a superficial knowledge of the pilot operating handbook for a flight review or an insurance checkout but an in-depth knowledge of the aircraft systems. Knowledge tests are usually generic. It would be unfair to ask questions about a particular type of aircraft, because a person taking the test may not have flown that type of aircraft. The same is true for the avionics installed in an aircraft. On instrument tests, the FAA uses National Aeronautical Charting Office (NACO) charts, whereas many airlines, such as Delta, use Jeppesen charts for their navigation utilities (Delta Corporate

Communications, 2013). Because the FAA does not ask questions about the specific type of aircraft a pilot may be flying, the type of avionics equipment installed in the aircraft, or even the type of charts the pilot uses for navigation, this can lead to gaps in determining a pilot's knowledge. An examiner should cover these areas during the oral portion of a practical test, but there are limits to the amount of material an examiner can cover. In a knowledge test, you cannot cover every facet of instrument flying in a 60-question test. Similarly, an examiner must cover every task and every area of operation, but he cannot cover all possible facets of aircraft systems, avionics, and navigation charts.

Procedural

Novice pilots normally respond to situations in a linear, problem-solving approach, according to Adams (1992). The pilot usually follows procedures he or she has learned and applies rules of thumb. The FAA (2009) advocated the DECIDE model, which is an acronym for Detect, Estimate, Choose, Identify, Do, and Evaluate. The expert, on the other hand, can alter his rules and create new rules and patterns, even though he may not have previously encountered a problem with these characteristics. He often seems to be able to solve problems at a rate that makes it appear intuitive.

Both military and commercial aviation are highly structured environments in which checklists have been developed for almost every situation. Shappell and Wiegmann (2003) pointed out that situations can be misdiagnosed or unrecognized or the wrong procedure can be applied. Procedural errors were found, by Wiegmann and Shappell (1997), to result in more minor accidents, whereas decision errors often resulted in major accidents.

Error

The brain functions as an error-detecting and error-correcting system, according to Campbell and Bagshaw (2002). Unfortunately, if it is given insufficient or misleading information, it can also be an error-generating system. To make good decisions, the brain must have adequate knowledge, skill, and experience, and must have reliable sensory inputs. It must also be rested and alert. If a person is tired or preoccupied or is too relaxed or insufficiently vigilant, he or she may be more likely to commit errors.

In a study of time-related human errors, it was found that 90% of these errors occurred during the preflight or taxi-out phase, as indicated in Figure 2.1. According to McElhatton and Drew (1993), incidents resulting from these errors were often found in the same or succeeding phase. They were often not caught on the same phase because these are high-workload periods, and the error was sometimes not caught until there was a reduction in workload. A graph of error detection would appear to be exponential, with half the errors being detected in the first minute and half in the second minute; thus, many of the errors were caught in the first few minutes after commission. Errors occurred more frequently when the pilot was under pressure due to maintaining a schedule. With combined distractions, schedules, and high workload operations, the pilot faced an increased chance of committing errors. The authors further categorized human error as cognitive, perceptual, or motor. The cognitive errors were those involving execution errors. Cognitive errors accounted for the largest percentage of errors, as shown in Figure 2.2.

When looking at aircraft accidents, Detwiler et al. (2006) classified errors as skill-based, decision, or perceptual errors; decision errors would equate to cognitive errors as described by McElhatton (above). Skill-based errors were doing errors—ones involving technique or memory and attention. The next errors were decision or thinking errors. These involved inability to make correct decisions on takeoff or landing or during fuel management. The last type of error, the perceptual error, resulted from inaccurate use of perceptual information. This can be caused by illusions or spatial disorientation and may cause a pilot to make errors in such things as judging distances and descent rates. Studies have been conducted to determine the phase during which flight accidents occurred. It was determined that 28% of accidents occurred during landing, 19% on takeoff, and 12% during approach (Trollip and Jensen, 1991). These are the three times when workload is highest for the pilot.

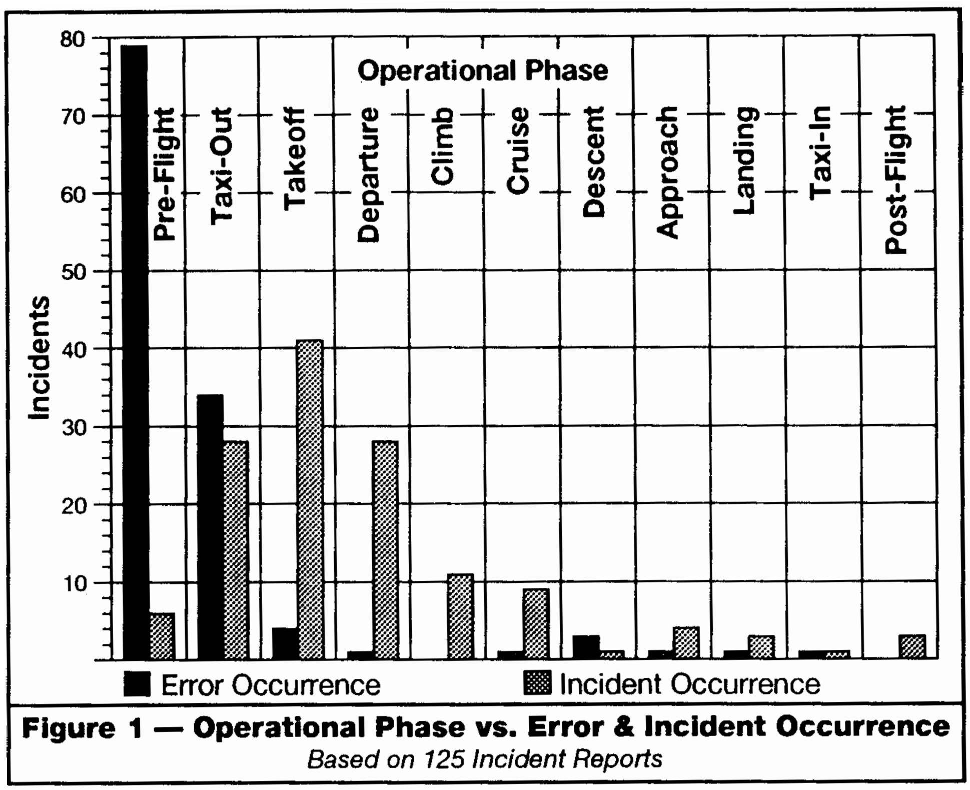

Figure 2.1 Phase of flight in which error occurred and the phase of flight in which an incident occurred as a result of the error. From " Time Pressure as a Causal Factor in Aviation Safety Incidents: The "Hurry Up" Syndrome" by J. McElhatton and C. Drew, 1993, Proceedings of the Seventh International Symposium on Aviation Psychology, April 22-29 1993, p. 274

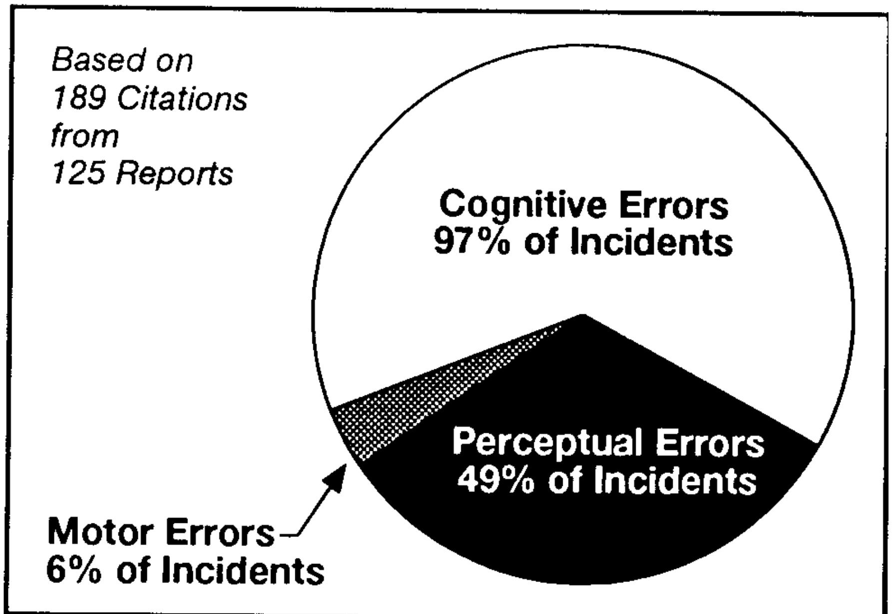

Figure 2 — Classification of Human Error

Figure 2.2 Classisfication of human error. From " Time Pressure as a Causal Factor in Aviation Safety Incidents: The "Hurry Up" Syndrome" by J. McElhatton and C. Drew, 1993, Proceedings of the Seventh International Symposium on Aviation Psychology, April 22-29 1993, p. 274

Amalberti and Wibaux (2000) pointed out that resources are limited, and pilots often have to take into account multiple risks. They want to avoid accidents, but they may be faced with information overload and need to guard against a degradation of their personal performance. Pilots may face a choice between what they would do in an ideal situation and what they are physically or mentally capable of doing. Errors, which can lead to accidents, often resulted.

Many have come to recognize that errors are inevitable. No matter how hard pilots try to prevent errors, they cannot. Boedigheimer (2010) presented a strategy in which a pilot first tries to prevent error. If that is not possible, then efforts should be made to try to trap the error before it occurs; if the error does occur, then the effects would be mitigated. By "trapping errors," the author meant that no negative consequences came from the error, or, for practical purposes, it did not exist.

Campbell and Bagshaw (2002) put it another way. Efforts should be made to minimize errors occurring as a result of the interactions among software, hardware, the environment, and liveware (the human element) and learning to live with the errors which will inevitably result.

Helmreich and Merritt (1998) developed five precepts for error management, which are paraphrased following.

- Human error is inevitable.
- There are limits on human performance.
- Humans make more errors when their limits are exceeded.
- Safety is a continuum from increased to deceased probability of accidents.
- High-risk organizations must develop a safety culture.

A promising area that may have been overlooked is error detection. Sarter and Alexander (2000) showed that a lot of effort had gone into developing error classification systems and designing error-tolerant systems. They also pointed out that much effort had gone into training and design, but relatively little had been done in creating error-detection processes.

Aviation Safety

Amalberti and Wibaux (2000) noted that errors are always going to occur. They will happen to people and at a time that one would never expect. Machines, on the other hand, fail in predictable times and manners that engineers can model and often easily repair. This contrasts with people, who can be unreliable. Engineers often apply more resources to improving machines because they think this is the best way they can control safety. There is a linkage between safety and training that may sometimes not be obvious. According to Maurino (1999), sometimes safety problems are answered with training changes. Training is also used as a fix when design flaws are found after an aircraft is in production.

In a study focused on the use of safety resources between pilots who had been involved in an accident and a group of pilots who had not, the only statistically significant difference that Rakovan, Wiggings, Jensen, and Hunter (1999) found was that the non-accident pilots hired a Certified Flight Instructor (CFI) more frequently than did the accident group. They also found that non-accident pilots were more likely to have attended a safety seminar than the accident group—a marginally significant statistic.

Knecht (2012) conducted another study about pilots involved or not involved in accidents. In this study, he looked at the source of training (either FAR Part 61 or 141) and who provided the check ride that awarded them their private pilot certificates. The three sources of certification were FAA inspectors, schools that had self-examining authority, and designated pilot examiners. He also looked at pilots with and without instrument ratings. He set up a 2x2x3x2 matrix. Several of the cells had fewer than five entries, so he could not use chi-square as his statistical tool. He used a log-linear analysis and had only one statistically significant finding: that those examined by FAA inspectors had fewer accidents. The author discounted this finding because it was based on only 22 events

Safety attitudes over the years have changed considerably. In 1982, Kinaszczuk, Dodge, and Mohler (1982) found that there were very few reports on pilot attitudes, because there was reluctance on the part of GA pilots to participate in studies of their attitudes and practices. He was able to conduct a study with the Dayton Flying Club and found a higher safety attitude among those pilots with more than 500 hours of flight time than among those with fewer than 500 hours. In 1999, Rakovan et al. (1999) conducted a study of GA pilots on the dissemination of safety information. They surveyed 6,000 pilots and had a response rate of 30.4%— very good response rate that showed a change in attitude towards surveys dealing with safety.

Safety cannot be achieved by rote rule-following. Dekker (2001) noted that situations change, and procedures have to be adapted. People have to be flexible and have to look at the context in which the rule is being applied. Studies have shown that many incidents occur because of errors made during the preflight or taxi-out phase of a flight. Because some of these errors occurred during weather decision making, Jensen, Guilkey, and Hunter (1998) conducted a study on personal weather minimums and regulatory guidelines for taking off. Personal minimums usually are more conservative than the existing regulations. The author was trying to determine if a two-hour safety seminar could influence pilots to use personal weather minimums and thus avoid takeoffs in weather that could cause pilot problems. He found that pilots considered the training helpful, would use personal minimums in their decision making, and would recommend the use of personal weather minimums to others.

Adams (1992) found it remarkable that, with all the effort that pilots make to learn meteorology, aircraft performance, systems, and navigation, little effort is put into learning how to organize and apply this substantial knowledge. PTS have improved in their requirement to include a scenario. Previously, many check rides were a series of maneuvers with little resemblance to an actual flight that the pilot would be conducting in the future. Wright (2012) wanted to see an additional test added to the knowledge and practical tests that would cover planning, risk management, and single-pilot resource management (SRM).

In recent years, Technically Advanced Aircraft (TAA) have been produced in large numbers. It is not possible to buy a new aircraft equipped with conventional round dials. It was thought that these aircraft—with their moving maps, which provide enhanced situational awareness—would lead to a decrease in the accident rate. Wright (2010a) pointed out that these aircraft have not led to the expected safety improvements. Using the same old training techniques, which rely on rote memorization and traditional training maneuvers, does not provide improved results. He believed that concentration on risk management and other higher order skills, in order to improve safety, is necessary. He faults the FAA; since the Colgan crash in 2009 in Clarence Center, New York, the agency has concentrated on airline pilot training instead of making the needed large scale curriculum changes in GA pilot training.

A new training method—FAA/Industry Training Standards (FITS)—has evolved, which has been successfully used with the introduction of TAA. Halleran and Wiggins (2010) cited three advantages of FITS. It utilizes scenario-based training (SBT), incorporates risk management, and stresses the limitation of TAA systems and systems management. A key concept of the FITS training is the development of self-assessment skills; it was hoped that these would help make a pilot a lifelong learner. It is hoped that the pilot, by the use of SBT and SRM, would move beyond the mastery of individual skills (as in the traditional method of flight instruction) toward safe mission accomplishment.

Decision Making

Flying is done in a dynamic environment with a variety of weather conditions and time and performance pressures. Pilots have to apply the knowledge they have gained to safely pilot aircraft (Beard and Geven, 2005). They have to make dozens of decisions, both before and during the flight, that affect flight safety. Studies have shown that there is no definite relationship between pilot demographics and decision-making tasks (Driskill et al., 1998). There is some evidence to show that older pilots, as well as those with more recent and total flight experience, tend to take fewer risks. This has been intuitively known for years, as illustrated in a saying attributed to E. Hamilton Lee, an early airmail pilot and a captain for United Airlines in 1949. "There are old pilots and bold pilots, but no old, bold pilots" (PILOT STORIES: No Old, Bold, Pilots section) (Smithsonian National Postal Museum, 2004).

A factor that can affect pilot decision making is stress. In a study of Navy personnel, sailors were given a questionnaire that asked about life events that they had recently experienced. Events were given a value called Life Change Units (LCU), ranging from 1 to 100; significant events, such as death of a spouse, received the highest value of 100. Trollip and Jensen (1991) reported that, of those with scores higher than 300, 79% reported illnesses and injuries within a year of the life crisis. Another fact reported was that there was a high correlation rate between high LCU scores and accident rates.

Murray (1997) cited the human information-processing system as contributing to the problem. Humans tend to have a single decision-making channel, and they process in a linear fashion. When people attempt to solve problems simultaneously, these tasks interfere with each other. Pilots can sometimes get so engrossed in a problem that they tune out other warnings of impending danger, or the brain get bored and does not pick up changes in aircraft performance or environmental conditions that could affect the safety of the flight. One of the most striking examples of fixation was Eastern Airlines Flight 401 on December 29, 1972. The crew noticed an unsafe indication in the landing gear; while attempting to solve the problem, they inadvertently disconnected the autopilot and crashed into the Everglades, resulting in the loss of 112 passengers and crew members (FAA, 2015). This was one of the accidents that lead to the introduction of CRM in the airlines.

Some pilots take a chance; if it works out, it reinforces their risk taking. Hise (2002) related a story of a lawyer who was killed on an IFR flight from Albany to Cape Cod. He crashed in the ocean on final approach to the runway. No one will know if he descended below the minimum descent altitude while looking for the runway, but he had done so previously, and it had paid off. This time, he was not as lucky.

A pilot often has a wide variety of choices when making decisions. Hopkins (1994) promoted pilots making what he called the Most Conservative Response Rule. In his opinion, if there is any doubt in a pilot's mind about the outcome of a decision he is going to make, he should always make the more conservation decision. There are always going to be conditions that are beyond the pilot's capability. A professional is able to recognize these situations and has the integrity to make sound decisions, even though they may not be the most popular decisions.

Pilot Demographics

While there have been many studies done on the demographics of those pilots involved in accidents, there have been few describing the overall demographics of the larger pilot population. Hunter (1995) undertook this work; he started with 20,000 pilot records and culminated with 6,735 usable returns to use in his 1995 report. This may give us a standard to distinguish the characteristics of the pilot population that has been involved in accidents from the general pilot population.

There have been some studies done to try to determine if pilot demographics have an effect on aviation safety. In a comparative analysis, Urban (1984) determined that:

> *In summary, based on the results of this study, pilots who possess comparatively high levels of education, espouse a liberal political ideology, are among the oldest siblings in their families or orientation, fly professionally, report comparatively more activity in nonflying aviation events, engage in relatively more aviation-related nonconformist*

behavior, and comparatively have not been involved in aviation for very long, appear to be more likely involved in aviation accidents. (p. 312)

In a survey concerning pilot decision making conducted by O'Hare and Chalmers (1999) in New Zealand, they were able to determine that exposure (among those surveyed) to hazardous events was common. He was able to compare these findings to data developed by Hunter (1995) in the United States. From the comparison, it appears that flying is flying. National boundaries or societal norms do not seem to affect pilot experience. In another study involved with VFR cross-country crashes, O'Hare and Douglas (2002) made some interesting discoveries. They analyzed crashes involving factors that were externally driven (EXD), such as engine failures, and compared them to crashes they called “In Flight Volitional” (IFV) (p. 366), which were mostly weather-related or due to loss of control. In this study, they found that the average age of the pilots in weather-related crashes was 37.8, while that of those in the externally driven group was 47. Some of the crash characteristics were also significant; the IFV crashes averaged 2,970 feet above sea level, while those in the EXD category averaged 150 feet above mean sea level. In terms of fatalities, there were twice as many fatalities in the IFV crashes as in the EXD crashes. In looking at the distances of the crash sites from the departure airfields, there was not much difference: total EXD crashes were 78.1 nautical miles, and IFV crashes were 72.9 nautical miles. If, however, you separate the weather-related and loss-of-control accidents, you find that weather-related accidents tended to happen farther away from the departure point (92.5 nautical miles) than the loss-of-control accidents (49.7 nautical miles). In another study, O'Hare, Wiggings, Batt, and Morrison (1994) looked at information-processing error versus the type of pilot certificate held. They found that, 96% of the time, student pilot error could be coded into one of six cognitive errors. They found the rate for commercial pilots to be 69% and holders of airline transport rating committed cognitive errors in 50% of their mishaps, which they thought to be high for their level of training and experience. They found that diagnostic and goal errors were more likely to result in fatal crashes. Their study indicated that errors committed early in the chain of events—where an incorrect goal was established or there was incorrect situational awareness—resulted in accidents more serious than those with an error later in the chain on how to do a particular procedure.

Of particular note are VFR pilots flying into IMC conditions. Goh and Wiegmann (2002) determined that, between the mid-1970s and the mid-1980s, 19% of the fatalities occurred because of this type error. They also noted that fatalities occurred in 72% of these incidents, whereas the rate of fatalities in all other accidents was 17%. They also observed that this was a worldwide problem, with similar results in Great Britain and New Zealand. In looking at the pilot ratings of those involved in these accidents, they found that many of the pilots in accidents with other causes had instrument ratings, while few in the VFR-to-IMC accident had instrument ratings. A chi square analysis found that ratings versus type of accident were significant.

In another survey conducted by Vail and Ekman (1986), it was found that men had a higher accident rate than women and that men were involved in a higher percentage of fatal and serious accidents. In looking at occupations, it was found that doctors, lawyers, sales representatives, farmers, and housewives had higher instances of accidents than other GA professions. So many doctors were killed piloting the Beech Bonanza that it earned the nickname the “doctor killer.” Pierce (2013) found that, during World War II, the accident rate for the Women’s Air Force Service Pilots (WASPs) was .05 fatal accidents per 1,000 flight hours, while that of the entire Army Air Force was .07 fatal accidents per 1,000 flight hours.

Hackworth, King, Cruz, Thomas, Roberts, Bates, and Moore (2007) noted some possible reasons for increased accident rates among GA pilots. These included a higher number of takeoffs and landings per hour, which are the most dangerous phases of flight. They also noted that the types of operations of GA pilots can be more dangerous. These can include agricultural flying, which is normally very close to the earth and can be near power and telephone lines, and instruction, which involves flying with students with very little experience. Overall, they concluded that GA pilots have less experience than commercial pilots, which could lead to their higher accident rate.

In a study of mid-air collisions (MAC), Taneja and Wiegmann (2001) found that the rate of MACs has remained fairly constant over the last few years. An interesting distribution of pilots involved in these accidents was found. Approximately 38% of the pilots involved had fewer than 1,000 hours of flight time. The other peak involved 25% of the pilots having more than 5,000 hours. The main cause of these types of accidents was attributed to inadequate look out by the pilot/pilots by the NTSB.

According to Rolfe (1972), no one was born with the ability to operate complex equipment. People must have sufficient knowledge and be trained to operate this equipment. GA accidents often point out that the cause of many accidents is pilots attempting maneuvers for which they have not been trained or that are too difficult for their experience level. He pointed to a study of 41 accidents; the pilot in 22 of them had fewer than 150 hours of flight time. He also made an observation about recent flight experience when he referred to a study of Royal Air Force pilots by Smith (1966), which showed that pilots with fewer than 100 hours of flight time in the preceding year had more accidents than the remainder of the pilot population. Smith also found that the highest accident rate was among 21- to 23-year-old pilots who were on their first operational assignments. He attributed this to their being freed from the constraints of the training environment and being allowed relative freedom in the operational environment. With pilots aged 23 to 29, he noted a decrease in the accident rate, which he attributed to an increase in experience and skill and the fact that pilots in this age group have good health and a high level of physical fitness.

In a study of air traffic controllers, Oakes, Ferris, Martocchio, Buckley, and Broach, (2001) attempted to use personality traits as a predictor of success in becoming a controller. The thought was that personality and motivation would be predictors of skill acquisition. They found that there was a positive correlation between four of the seven tested personality traits and successful controllers. The military has used tests for years in the pilot selection process. While selection testing may not be practical in GA—because anyone with the money and the desire can take pilot lessons—it may be a tool that could identify those who may have difficulty acquiring the necessary skills and thus possibly perform at a lower level.

Knowledge Testing

AFS 630 is the FAA agency responsible for creating and monitoring the administration of knowledge tests. Two companies, PSI/Lasergrade and Computer Assisted Testing Services (CATS), are contracted by the FAA to administer these tests., which vary in the number of questions and the length of time allotted for completion. They generally consist of four types of questions—recall, define, calculate, and interpret—all of which are in a multiple-choice format. While the first two types of questions are answerable by rote memorization, the last two should require the applicant to understand and apply the material. There has been a difference of opinion concerning the publishing of the question banks from which questions are drawn for the tests. In

2003, the FAA limited the information released about questions to the form of the question but did not use the same data. (Casner et al., 2004)

In 2012, the ARC questioned the reliability and validity of the questions in the current knowledge bank. The committee proposed that all questions be returned to the public domain (Hennig, 2012). The FAA rejected this recommendation. The agency's concern is that students will study the questions to pass the test and not understand the material. The FAA has had concerns for some time about pilots memorizing answers to test questions. With the advent of computer testing, the agency has been able to track the amount of time a pilot takes to answer individual questions (Casner et al., 2004). It is obvious, as shown in Table 2.1, that some pilots complete the test in less time than a normal person needs to merely read the questions. The unknown factor is: does the pilot really understand the material, or is he or she memorizing the answers in order to facilitate taking the test?

Table 2.1
Test Completion Times

Test	Ave. Completion Time	Minimum Time
Private Pilot - Airplane	73	2
Instrument Rating - Airplane	79	2
Commercial Pilot – Airplane	91	11

Note. Time is in minutes. From " FAA Pilot Knowledge Test: Learning or Rote Memorization" by S. Casner, K. Jones, A. Puentes and H. Irani, 2004, NASA/TM-2004-212814, p. 3

There is some concern in the aviation industry that some of the questions in the knowledge test bank are dated and that questions involving newer technologies, such as TAA, are slow to be added. Goyer (2010) was particularly concerned that there are far too few questions actually related to what pilots are doing when they are flying. He also believed that many of the flight-planning questions have answers that are too close together. In real-life flight planning, the answer is constantly changing; if ATC gives a slower than expected climb to your cruise altitude or if the winds are not as forecast, then the time enroute can vary more than the difference in the FAA answers. He felt that the main purpose of close answers is to get many of the test takers to choose the incorrect answer. This provides a bell curve and the appearance that the test is measuring the knowledge of those taking it. Wright (2010b) was concerned that neither the knowledge test bank nor the PTS reflects the latest FAA doctrine and guidance. Those who would like to contribute questions to the FAA knowledge test bank may do so by email them to afs630comments@faa.gov.

To correct many of the problems with knowledge testing and the special-emphasis area of the PTS, AFS 630 recently published a draft Private Pilot Aircrew Certification Standards (Federal Aviation Administration, TBD). Each task has sections for objective, knowledge, skills, and risk management. This now unites the knowledge that is needed for a maneuver, the skills necessary to perform the maneuver, and the risk assessment and risk mitigation that may be necessary prior and during the maneuver.

Fixed Versus Adaptive Testing

The current FAA knowledge tests are fixed form; the number of test questions varies from 20, for some helicopter instructor add on ratings, to 125, for the airline transport rating and military competency instructor rating (Federal Aviation Administration, 2014d). Ortner and Caspers (2011) believed that computerized adaptive testing (CAT) is one of the most outstanding technological innovations in the last century. They noted that measurement error is noticeably reduced, even when using fewer test items. Weiss and Betz (1973) found that the adaptive format was more reliable and valid than conventional testing. They also found promise that it would be fairer for minority groups. There are a couple of features that students do not like when taking adaptive tests. Students are not able to go back and review previous questions, and they cannot skip questions. If a student were able to go back and change previous answers, he or she may be able to game the test. If the next question were harder, the student could assume the previous question was correct; if it were easier, that the previous answer was incorrect. If you were allowed to skip questions, the computer would not know the level of difficulty that the next question should address.

Foreign Testing

In Europe, the Joint Aviation Authorities (JAA) requires as many as seven tests for the private pilot certificate (Casner et al., 2004). Some of these tests may consist of as few as 15 questions. As in the United States, all questions are multiple-choice format. The JAA does not publish a list of currently used questions, but it does publish old questions and similar questions. There is a difference in the passing score that an applicant must attain. In Europe, the passing score is 75%, but the applicant must attain that score on each of the tests (Government Accountability Office 2011). The Australian Civil Aviation Safety Authority (CASA) uses a system similar to the JAA and does not publish current questions. India uses a system of five tests for the private pilot certificate. The questions are multiple-choice and are not published.

Other Discipline Testing

Many other professions utilize knowledge testing in their certification processes. The National Council of State Boards of Nursing (NCSBN) has much tighter controls over test information disclosure than does the FAA (AFS 630, 2012). The council copyrights the test and makes cheating a criminal offense. The NCSBN provides neither immediate feedback nor detailed information about questions missed. The NSCBN uses a computer-adaptive format, whereas the FAA uses a fixed form.

The Airmen Testing Standards and Training Aviation Rulemaking Committee looked at the testing procedures of several other occupations when they were making their report to the FAA. The occupations included: Doctor, Nurse, Teacher, Accountant, Air Traffic Control, and Coast Guard. Most of them use multiple-choice questions in a computer format. None of them have public databases. The ARC recommendations could be influenced by the fact that several members of the committee are in the business of producing training material for students preparing for knowledge tests.

Practical Testing

Practical tests are conducted by FAA inspectors and designated examiners; in some cases, schools have self-examining authority, so their final progress check ride counts toward the issuance of the certificate sought. In all cases, the check ride is given in accordance with PTS. After successfully completing his or her knowledge test, a pilot fills out an Integrated Airman Certification and Rating Application (IACRA), which is approved by the flight instructor before an appointment for a practical test can be scheduled. The PTS, according to Wright (2012), describe the maneuvers to be conducted and establish an objective measurable minimum standard for each maneuver. While the PTS have been improved over the years to include risk management and other single-pilot resource management (SRM) skills, Wright did not feel that the current construct of the test was adequate for testing these skills. He proposed a three-phase test for a certificate. The knowledge and practical tests would be similar to the current tests but would include another test that evaluates a pilot’s planning, risk-assessment, and SRM skills prior to the practical test. It would be administered in much the same way as the knowledge test.

There is some criticism of the PTS in that they rely on performing specific maneuvers instead of testing the higher-order skills required of a pilot during an actual flight. Wright (2011) wanted to see the expanded use of scenarios to test higher-order skills and a reduction on the emphasis of specific maneuvers.

Recertification

Pilots have a certain level of knowledge and proficiency when they are issued pilot certificates. However, lack of retention and the forgetting curve affect the amount of knowledge a pilot remembers. The FAA attempts to ensure that a pilot is safe by requiring him or her to complete a flight review, consisting of one hour of flight and one hour of ground training every 24 calendar months (Federal Aviation Administration, 2014a). The ground training must cover the flight rules in FAR Part 91 and those maneuvers and procedures necessary for the safe exercise of pilot privileges. A pilot cannot fail this review. If he or she does not satisfy the expectations of the flight instructor conducting the review, the instructor will make an annotation in the pilot's log book that he or she received training, but there is no mention that he or she failed the review. The pilot can study and practice and return to the original instructor or seek another instructor without any training to see if that other instructor will sign the review. In other industries, such as medicine, human resources personnel, auditors, engineers, and quality technicians often go through a process of recertification that can consist of examinations or continuing education as evidence of continued competency. According to Bashook and Parboosingh (1998), a snapshot assessment every seven to 10 years is a crude way of measuring the medical skills of doctors. They wanted to see a continuous assessment of their practice and evidence of continual learning in their field.

The Institute of Electrical and Electronics Engineers (IEEE) sees value in reaccreditation, because, in fast-changing fields such as wireless communications, technologies can change rapidly, and skills can become outdated (Frantz, 2011). One way the IEEE sees of recertifying is to retake the examination, which is continually changed to keep up with changing technologies. The test one takes for recertification would be nothing like the test taken five years earlier. The test is only a portion of the recertification process. Taking courses, attending professional workshops, authoring papers, and speaking at educational sessions are all activities that could be considered for professional development points and presented to a panel of wireless professionals. Recertification could be awarded for the accumulation of points in a variety of categories.

The American Composites Manufacturers Association has a fairly standard certification process. The applicant must take a 100-question, multiple-choice examination that has to be taken at a designated computer center or as a written examination with an approved proctor. The certification is good for a three-year period. The recertification process consists of reviewing a recertification workbook and taking an open-book examination (American Composites Manufacturers Association, 2015).

The only aviation assessment that appears close to the Bashook and Parboosingh or IEEE model is that used by Master Instructors LLC to designate masters in the educator, flight instructor, and ground instructor designations (Hill and Hill, 2014). This designation is required to be renewed biannually. It requires a minimum of 500 hours of activity in five categories, with a corresponding minimum number of hours in each category. A panel of peer Master Instructors evaluates the activities submitted. Educator is the main category, with a minimum of 250 hours of activity required. The other categories are service, media, continuing education, and participant; each category requires a minimum of 30 hours. This program ensures that designees remain active as both teachers and learners.

The WINGS program has previously been discussed as a voluntary program for GA pilots to help them build an educational curriculum that is tailored to each pilot's individual flight requirements. It consists of both knowledge and flight activities. There are 22,273 pilots (Federal Aviation Administration, 2014c) who have earned at least one phase of the WINGS program. It would seem possible that this well-established program could serve as a basis for an expanded program that could improve upon the one hour of ground and one hour of flight that are required every 24 calendar months in our current flight review program. The WINGS program includes a history of participation that is maintained in a computer database; it provides recommended activities, based on the pilot's profile, for each level of the WINGS program. There are seminars that are scheduled by FAASTeam members and online training that can be taken. This includes not only FAA training but also training offered by the Aircraft Owners and Pilots Association (AOPA) Air Safety Institute, Civil Air Patrol, and many other organizations that have courses, seminars, and flight training eligible for inclusion in the program.

The segment of the pilot population with the fewest accidents is professional airline pilots. The last accident by a U.S.-scheduled passenger-carrying airliner was an accident by a regional airline, Colgan, in February 2009 in Clarence Center, New York. In this accident, 50 people lost their lives. Since that time, there was a Boeing 777 crash in San Francisco involving a foreign carrier, Asiana Airlines, and an Airbus 300 crash in Birmingham, Alabama, by United Parcel Service, a cargo carrier.

GA pilots cannot be held to the same training frequency, duration, or intensity as airline pilots; however, looking at airline pilot training may offer some clues to their remarkable safety record. Airlines fly and train according to the regulation set forth by the FAA (2014b) Part 121, Operating Requirements: Domestic, Flag, and Supplemental Operations. All flight-crew members are required to undergo initial, transition, upgrade, and recurrent training. Recurrent training is required annually for all crewmembers, and pilots in command are required to complete an additional training period within the previous six months. For pilots of group II or turbojet aircraft, this training has to consist of a minimum of 25 hours. There is also a requirement for demonstrating proficiency in a series of maneuvers in a flight simulator. These simulators are exact replicas of actual aircraft. Levels C and D simulators must have a motion platform with 6 degrees of freedom and a visual depiction of 150 degrees horizontally and 30 degrees vertically (FAA, 1991).

Maintenance

A study of maintenance-related accidents that occurred between 1988 and 1997 found that 1,474 or 7.1% of all GA accidents had at least one maintenance-related error as a cause of the accident (Goldman, Fiedler, and King, 2002). Because of limitations on NTSB coding of accidents, this is probably conservative, as other military and GA studies have shown accident rates up to 20%. On average, 18% of the maintenance accidents resulted in fatalities. Pilot and maintenance errors account for most accidents. Once an accident is determined to be maintenance-related, little further investigation is conducted. Many of these accidents are probably caused by human error. The most significant type of maintenance error appears to be installation errors. The two types that resulted in the most GA deaths were reverse installation and using the wrong part.

The aviation maintenance technician (AMT) is primarily concerned with the safety of the aircraft on which he is working (Patankar, Brown, and Treadwell, 2005). Once an AMT gets promoted to a managerial position, the focus often switches from safety to profit. Eliminating pilot errors and maintenance mechanic errors would eliminate most accidents.

Future Systems

To improve aviation safety, the National Aeronautics and Space Administration (NASA) has established the Small Aircraft Transportation Systems (SATS) to increase the safety of single-pilot GA operations. Snyder, Sokoloff, and Bearden (2003) envisioned a co-pilot in a box for GA pilots. These systems would improve pilot situational awareness by performing some of the information-management and communications tasks now done by co-pilots. They would also be able to conduct error monitoring, by checking fuel status and monitoring for actions, such as going below the minimum descent altitude on an approach, that would violate regulations. Such a system would also be able to do environmental monitoring of such things as airport status, changes in the weather, and airspace availability. With the increases in avionics technology, pilots can become saturated with data. This new technology would give us the possibility of filtering this information to present to the pilot the items that are more important for situational awareness.

In a GA study conducted by Saleem (1996), three pilots conducted an IFR flight to the Roanoke airport, utilizing a safety pilot and an ethnographic observer. On all three flights, the pilots made mistakes. After studying the situations, Saleem came up with the following as possible solutions. It may be possible to filter communications traffic heard by the pilot. Few of the many radio calls on a frequency assigned to a pilot actually are to or from him, yet he must listen to them all. Displays to reduce the burden on working memory could show heading, altitudes, and other information that a pilot must know, remember, or take the time to write on a knee board. The last suggestion is one that the FAA is actively pursuing. It involves auditory and visual traffic information displays to help pilots locate and avoid other airplanes. Air traffic controllers currently provide this service to IFR traffic and upon request (and depending on how busy the controller is) to VFR traffic. Some aircraft are equipped with traffic-collision avoidance systems, which sense the emission from the other aircraft's transponder and produce auditory and visual representations of the other aircraft's position. At the current time, not all aircraft are required to have a transponder, so the system will not report these aircraft. In the future, all aircraft with be required to have a system called Automatic Dependent Surveillance-Broadcast (ADS-B), which will assist pilots in all phases of flight but especially in busy terminal areas.

Summary

GA accidents continue to occur at a rate far in excess of airline accidents. Within the GA statistics, pilots who are subject to regular recurrent training have fewer accidents than pilots who meet the minimum requirement of the regulations, which is a flight review every 24 months. Women are safer pilots than men; men aged 35 to 39 who have 100 to 500 flight hours and are on personal flights in VMC conditions are most likely to be involved in accidents (Trollip and Jensen, 1991). The most likely causes are a lack of knowledge, complacency, and inadequate flight training. What is not known—and is hoped to be a contribution to the body of knowledge by this paper—is determination of a link between knowledge test results and accident statistics.

Chapter 3 - Methodology

The purpose of this study was to see if there is any relation between accidents and FAA knowledge test results. Quantitative methods were used in the study. Data were collected in an Excel spreadsheet, in which simple calculations were made; data were then loaded in SPSS for more involved calculations. Data were obtained for this study from AFS 630, which obtained the accident data from the NTSB. AFS 630 maintains the database of knowledge test questions and the results of all knowledge tests administered by their two contractors. AFS 630 took the data from the NTSB and located the test results for the pilots who had been involved in fatal accidents. Once the test scores were married with the accident data, the results were de-identified.

Research Questions

This research was guided by the following questions, in relationship to the causes of aviation accidents. How can the number of GA fatal accidents be reduced?

1. Do pilots involved in fatal accidents due to human error score lower on FAA knowledge tests than pilots involved in fatal accidents due to mechanical failures?
2. Do pilots involved in fatal accidents due to human error spend less time per question on FAA knowledge tests than pilots involved in fatal accidents due to mechanical failures?
3. Do pilots involved in fatal accidents due to human factors errors spend less time on recall questions than pilots involved in maintenance factors accidents.
4. Do pilots involved in fatal accidents due to maintenance factor spend more time on calculate questions than pilots involved in human factor error accidents.

Accident data

Accident data were drawn from the NTSB database for the period 1/1/2009 to 12/31/2011. Of the 802 fatalities, 139 had taken a knowledge test between 1/1/2008 and their deaths. Data prior to 1/1/2008 were excluded to minimize changes due to lack of retention or changes in experience level. The accident reports include FAA remarks, NTSB remarks, and the NTSB findings. The NTSB findings list things that were determined to be a cause or a factor in the accident. Based on previous studies, it was anticipated that, in approximately 10% of the accidents, the NTSB (or, in some cases where the investigation was delegated to the FAA), the causes of the accidents would not be able to be determined. In the remaining cases, approximately 75% of the accidents would be due to human error, and approximately 25% of the accidents would be caused by mechanical errors. Based on the sample size, this should have provided approximately 14 accidents of undetermined origin, 94 accidents caused by human error, and 31 caused by maintenance problems.

Knowledge Test Data

Of the 139 pilots who had taken knowledge tests, some had taken multiple exams. The examination results provided by AFS 630 included the type of examination, date, and score. For each question, the following information was provided: learning statement number and verbal description, which indicates if it was a recall, define, calculate, or interpret type of question; whether the question was correct or incorrect; and the number of seconds that were spent answering the question. Because each pilot had to pass the appropriate knowledge test to obtain his rating, scores fell mostly between 70 and 100. There were a few scores below 70, in which pilots failed a test for the next higher rating but were still able to fly at their then-current ratings. The questions that were of more interest were the calculate or interpret questions, which are skill questions. These are questions that should take longer to answer than knowledge questions, which can rely on rote memorization. Looking at the average time to answer skill questions may provide an indication if a particular pilot memorized the answers. Table 3.1 summarizes the variables that were used.

Table 3.1

Variables

Variable	Variable Description	Min	Max	Mean	Std. Dev.
Date of Birth		14 Feb1927	4 Nov 1992		
Date of Event		30 Jan 2009	20 Dec 2011		
Type of Accident	Either Human Factors or Maintenance	11 Maintenance Related	106 Human Error Related		
Type of Test	There were 21 total types of tests	7 types of tests with 1 participant	IRA 26		
Date of Test		9 Jan 2008	29 Jul 2011		
Score	Total Score for the test	52	98	84.38	8.776
Calculate Time	Average time to complete each calculate question	27	374	147.55	68.488
Recall Time	Average time to complete each recall question	15	126	48.15	21.892

Total Time	**Total time for test. Max 3 hrs or 10,800 seconds**	**760**	**10745**	**5211.69**	**2185.432**
Calculate	**Score on Calculate questions**	**45**	**100**	**81.62**	**13.19.6**
Recall	**Score on Recall questions**	**51**	**100**	**85.37**	**9.216**
Age at time of event		**18**	**84**	**42.42**	**14.236**

Control Procedures

Because this was a retrospective causal-comparative study, it was not possible to use randomization to select an experimental and control group. The two groups that were used can more accurately be described as comparison groups. A method of equating the two groups that could be used is matching. Because it is known from previous studies that there are three to four times more pilot-error accidents than maintenance mishaps, using this technique reduced the size of the sample considerably. In this study, it was more feasible to compare two subgroups of the accident population: those where pilot error was the cause and those with causes originating from mechanical events. There are a small number of accidents for which the NTSB could not determine a cause. These were not used in the study.

This proposal was submitted to the Institutional Review Board (IRB) for Kansas State University, which determined that this was a Non-Research Application. It does not require review by the Committee for Research Involving Human Subjects.

Data Analysis

To analyze the data, a variety of descriptive and inferential statistics were used. Because test scores should vary between 70 and 100 and test times cannot exceed 3 hours, the mean was used to measure central tendency. Some pilots may have failed a knowledge test. In that case, the pilot rating held may be different from the last test result. The median might be more appropriate if a wider distribution of scores or times were expected. The mode would be an unstable measure.

The pilots in the study are all deceased GA pilots. Their demographic data (age and total flight times) were compared to previous studies to give an indication of whether the sample was representative of the pilot population. The FAA publishes national averages of pilot ages by type certificate. The average age of the deceased pilots was compared with the average from the FAA data. The group was then divided, as indicated by the NTSB, into those involving pilot error and those that had mechanical failures.

The data were examined to determine if there were any outliers. Care was taken to remove the outlier only if it was not from the intended population sample. If the distribution appeared to be skewed by the outliers, it may have been necessary to apply transformations to the data. This also would have been necessary if the Levene's test for homogeneity of variance was significant. As a last resort, the score could be changed using one of several methods. Only test results related to flight were used. Any accident, in which the pilot's last test was maintenance or air traffic control, was removed.

Variance was used to determine the standard deviation, which is the most stable measure of variability. It was used to see if scores were clustered around the mean. Assuming that knowledge test results are normally distributed allowed the more powerful parametric testing to be used. Sample sizes between human-factor errors and maintenance failures were expected to be different. These different samples were not expected to cause any problems, due to the robustness of the ANOVA.

For the first research question, an ANOVA was used on the test scores to see if there were significant differences. For a group of data, Field (2009) suggested that use of an F-statistic or F-ratio would be appropriate. ANOVA is an omnibus test, so it is an overall effect. It does not tell specifically what the effect is. If the test is significant, it can be concluded that the means are significantly different. The ANOVA itself will not distinguish between groups. If there were a control group, planned contracts could be done. Since there were no control groups, post hoc tests were done. There were about 18 post hoc tests available in SPSS. Some were eliminated, based on sample sizes. The number of samples in each group and the differences in sample sizes among groups helped in the selection of post hoc tests. If a test was too conservative, a small difference would have been rejected when it was, in fact, meaningful. Because samples sizes were expected to be very different, Gabriel's and Hochberg's GT2 was appropriate for use. If the population variances were different, the Games-Howell procedure was run. In the options box, Brown-Forsythe and Welch was selected. These tests were useful if the homogeneity of variance assumption was broken, which was known by looking at Levene's statistic. If the test were nonsignificant—i.e., $p > .05$, then the variances were roughly equal.

For the second question, the time spent on knowledge test questions, a slightly different procedure was followed than that used for test scores. Different FAA knowledge tests have varying numbers of questions, and the ratio between the numbers of calculate and recall questions can vary from test to test. For this question, the calculate time and recall time per question were used. An ANOVA was used to compare the calculate and recall times per question of the groups. Outliers were checked, since it is known that the longest FAA test is three hours. Because the sample sizes in this question were almost identical to question one, the same post hoc tests and option tests were run for this question as for the knowledge test scores in question one.

The next issue was to see if there is any relation between questions answered correctly or incorrectly and the time taken to answer the questions. Using two groups, those with pilot errors and those with mechanical failures, a Factorial Anova were performed. Different FAA tests have different numbers of questions, so it was necessary to look at the percentage of questions answered correctly and not the number of questions answered correctly or incorrectly. Looking at the times for responding to both knowledge questions and calculate questions was performed. Scores were selected as the dependent variable and time and type of accident as the fixed factors. Results began with descriptive statistics showing means and standard deviations. Next, Levene's test for equality of error variances was examined for significance. If the value of the Levene's test was significant ($p < .05$), the variances were significantly different. If the value of the Levene's test was not significant ($p > .05$), there was not sufficient evidence to reject the null hypothesis. The main ANOVA table was then examined to see if there was a significant main effect for time. The significance value had to be less than .05. Contrast was also examined. Again, the significance had to be less than .05. A post hoc analysis was run and interpreted as if it were a one-way ANOVA.

Chapter 4 - Results of the Study

The data in this chapter are presented to answer the four main research questions. Initial results include overall descriptive charts and statistics, followed by the four research questions. In a study by Trollip and Jensen (1991), they found that, in the late 1980s, the profile for pilots most likely to have an accident was 35 to 39 years old. The results of this study showed a somewhat older group, with a mean of 44. Looking at the statistics published by the FAA for the years 2002 to 2013, the age of our accident population is virtually the same as the average age of active pilots, which lends support to the fact that our sample population was representative of the general pilot population. There has been a decrease in the number of private pilot certificates issued, from a high of more than 50,000 in 1979 to about 25,000 in 2013. While we generally think that the pilot population is aging, the statistics show that the average age of a pilot has remained virtually the same. In 2005, a new category of pilot certificate, the sport pilot certificate, was instituted. This certificate attracts a group of pilots who are older than the average. The statistics do show a slight decline in the average age of student pilots. Since 2007, airline pilots have been allowed to continue flying past age 60 and until age 65. This may have had the effect of increasing the average age of pilots holding an airline transport pilot (ATP) rating. It is interesting to note that, while the average age of a student pilot has gone down by two years, the average age of a private pilot has increased by two years. There has been a marked decrease in the average age of a recreational pilot, from 51 to 44 years old. Recreational pilots, however, make up a miniscule percentage of the overall pilot population. In 2013, there were only 238 recreational pilots among the almost 600,000 active airmen certificates held.

Table 4.1

Age Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Age at time of death	117	18	84	42.42	14.236

Table 4.2

Average Age of Active Pilots by Catagory

			Type of Pilot Certificates				
Calendar Year	Total 1/	Student	Sport	Recreational	Private 2/	Commercial 2/	Airline Transport 2/
2013	44.8	31.5	55.2	44.8	48.5	45.4	49.7
2012	44.7	31.5	54.7	47.8	48.3	44.8	49.9
2011	44.4	31.4	54.4	48.8	47.9	44.4	49.7
2010	44.2	31.4	53.8	50.8	47.6	44.2	49.4
2009	45.3	33.5	53.5	50.4	47.1	44.2	48.9
2008	45.1	33.6	53.2	50.1	46.9	44.8	48.5
2007	45.7	34.0	52.9	52.4	48.0	46.1	48.3
2006	45.6	34.4	52.9	51.5	47.7	46.1	48.1
2005	45.5	34.6	53.2	50.9	47.4	46.0	47.8
2004	45.1	34.2	N/A	51.3	47.0	45.9	47.5
2003	44.7	34.0	N/A	51.5	46.5	45.6	47.0
2002	44.4	33.7	N/A	51.0	46.2	45.5	46.6

Note. As of DECEMBER 31, 2013.

1/ Includes helicopter (only) and glider (only).
2/ Includes pilots with an airplane and/or a helicopter and/or a glider and/or a gyroplane certificate.
Pilots with multiple ratings will be reported under highest rating. For example a pilot with a private helicopter and commercial airplane certificates will be reported in the commercial category.
N/A Not available. Sport certificate first issued in 2005.

From "US Civil Airmen Statistics" by the Federal Aviation Administration, downloaded July 10, 2014, from http://www.faa.gov/data_research/aviation_data_statistics/civil_airmen_statistics/2013/

The average score on the FAA knowledge test was found to be 84, with a range from 52 to 98. The FAA passing score for the knowledge test is 70%. There were five pilots who had failed a knowledge test but were still flying. This amounted to 4.3% of the 117 pilots involved in human-error or maintenance-related accidents. All five pilots with failing scores were involved in human-error type accidents. The length of time from their failed tests until their fatal accidents varied from 8 days to 30 months. In looking at the means for the most common types of tests, the following results were obtained: PVT, 85%; IRA, 84%; and CAX, 83%. These scores are all very close to the mean of all tests; if the scores are normally distributed, between 70 and 100, they are close to what you would expect for the mean.

Table 4.3

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Score Test	**117**	**52**	**98**	**84.38**	**8.776**

Looking at the mean scores by accident type, the pilots involved in human-error accidents had lower test scores, with a mean of 84%, than those involved in maintenance-related accidents, with a mean of 87%. The mean for undetermined accidents fell between the means for human error and maintenance factors—which is understandable, since those accidents could result from a combination of human factors and maintenance factors. The FAA or the NTSB may have been unable to determine the cause of these accidents for a variety of reasons. There may have been no survivors or witnesses, the wreckage may not have been recovered or may have been too badly destroyed by fire, or, in some cases, there just were not conclusive indications of what caused the accident. In looking at the standard deviations, the smallest was that of the undetermined group. Their scores were closer to their mean than either the human-error or maintenance-factors groups.

Table 4.4

Type Accident Statistics

Type Accident	Mean	N	Std. Deviation
1	84.16	106	8.880
2	86.55	11	7.725
3	84.50	16	5.774
Total	84.40	133	8.454

Note. Human Error=1; Maintenance Factor=2; Not Determined=3.

In looking at the percentage of calculate questions answered correctly versus recall questions, pilots were less successful in calculate questions (81.5 %) than they were with recall questions (85.3%). There is a difference in the minimum scores of the calculate questions (45) and the recall questions (51). There is a difference in the number of pilots answering calculate questions versus recall questions because one military competency test had only recall questions. The calculate question had a wider range, and the standard deviation is larger, suggesting that the scores were more spread out.

Table 4.5

Calculate Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Calculate %	132	45	100	81.50	12.758

Table 4.6

Recall Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Recall %	133	51	100	85.36	8.866

Research Question 1

For the first research question—looking at knowledge test scores and the relative rank of those involved in human-factors related accidents as opposed to maintenance-related ones, or those of unknown origin—the results of Levene's test and the results of an ANOVA were examined. For the Levene's test $F(2, 130) = .120$, $\rho > .05$. The results of our ANOVA show that $\rho > .05$.

Table 4.7

Test of Homogeneity of Variances

Levene Statistic	df1	df2	Sig.
2.155	2	130	.120

Table 4.8

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	56.879	2	28.439	.394	.675
Within Groups	9377.001	130	72.131		
Total	9433.880	132			

Since there were no significant differences, post hoc testing was not warranted. Again, all of the significance values are greater than .05. As such, there is no significant difference in the knowledge test scores between those involved in human factors and those in maintenance-factor accidents.

Research Question 2

Research question two looked at the time the pilots took to answer questions. Tests vary in the length of time allowed for completion, so looking at the average time to complete the test did not provide any useful data, because the types of examinations taken by the pilots probably differed between the types of accidents. Looking at the mean number of seconds it took pilots to answer calculate and recall questions was of interest. Pilots took an average of 152 seconds to answer calculate questions versus 48 seconds to answer recall questions, which was expected. What was unforeseen was that pilots involved in human-error accidents took longer on both calculate question (149 seconds to 136 seconds) for maintenance-related accidents and for recall questions (49 seconds to 41 seconds). It was thought that those involved in human-factors accidents would be rote learners and would respond to questions more rapidly than those involved in maintenance-factor accidents.

Table 4.9

Calculate Time Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Calculate time	**132**	**27**	**374**	**151.74**	**70.640**

Note. Time in seconds.

Table 4.10

Recall Time Descriptive Statistic

	N	Minimum	Maximum	Mean	Std. Deviation
Recall Time	**133**	**15**	**126**	**48.00**	**21.222**

Note. Time in seconds.

Table 4.11

Calculate Time Descriptive Statistics

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum
1	105	148.75	70.400	6.870	135.13	162.38	27
2	11	136.09	47.302	14.262	104.31	167.87	62
3	16	182.13	80.611	20.153	139.17	225.08	51
Total	132	151.74	70.640	6.148	139.58	163.91	27

Note. Human Error=1; Maintenance Factor=2; Not Determined=3.

Table 4.12

Recall Time Descriptive Statistics

	N	Mean	Std. Deviation	Std. Error	Lower Bound a	Upper Bound	Minimum
1	106	48.85	22.261	2.162	44.56	53.14	15
2	11	41.45	17.369	5.237	29.79	53.12	21
3	16	46.88	15.987	3.997	38.36	55.39	17
Total	133	48.00	21.222	1.840	44.36	51.64	15

Note. Human Error=1; Maintenance Factor=2; Not Determined=3. a 95% Confidence Interval for Mean

For the Levene's test $F(2, 130) = .181$, $\rho > .05$. There was no significant difference in the recall or calculate times between our groups, as shown with the results of our ANOVA. Again Bonferroni was not needed. It is interesting that the not-determined group fell between the human-error and maintenance-factor groups (in research question one) in looking at their knowledge test scores. In research question two, they took considerably longer in answering calculate questions than in either the human-error or maintenance-factor groups. In looking at the recall times, they once again fell between the groups. Again, the difference in N between calculate time and recall time was due to a knowledge test having no calculate questions.

Table 4.13

Test of Homogeneity of Variances Total Time

Levene Statistic	df1	df2	Sig.
1.732	2	130	.181

Table 4.14

ANOVA Total Time

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7164003.796	2	3582001.898	.689	.504
Within Groups	676063527.723	130	5200488.675		
Total	683227531.519	132			

Research Question 3

Research question three looked to see if there was any relation between questions that could be answered by rote memorization and the time taken to answer the questions. In the descriptive statistics, the mean for the human-error accident for recall was 85%, while the mean for the maintenance-factor group was 89%. This difference was significant when we looked at the Levene's test. In looking at the averages by times within each group, they share the common pattern that they scored highest on the median times. The longer or shorter the time they spent on a question, the lower the average score they received.

Table 4.15

Dependent Variable: Recall %

Type Accident	recall time recoded1	Mean	Std. Deviation	N
1	.00	85.71	8.115	34
	1.00	86.13	7.550	39
	2.00	82.76	11.953	33
	Total	84.94	9.332	106
2	.00	88.80	10.826	5
	1.00	90.50	4.950	2
	2.00	89.75	1.708	4
	Total	89.45	7.118	11
Total	.00	86.10	8.404	39
	1.00	86.34	7.462	41
	2.00	83.51	11.493	37
	Total	85.37	9.216	117

Note. Human Error=1; Maintenance Factor=2.

For the Levene's test, $F(5, 111) = .003$ $\rho < .05$. This shows that there was a significant difference between group means.

Table 4.16

Levene's Test of Equality of Error Dependent Variable: Recall %

F	df1	df2	Sig.
3.840	5	111	.003

Note. Variances tests the null hypothesis that the error variance of the dependent variable is equal across groups.

To exam this difference, a factorial ANOVA was set up, using the recall percentage as the dependent variable; the type of accident and the recall time, transformed into a nominal numeric input, were independent variables. The resulting test of between-subjects effects showed $\rho > .05$.

Figure 4.1 shows that both the human-error and maintenance-factors pilots scored best when they took a median amount of time to answer questions. If they took too short or too long a period of time to answer the questions, their scores were lower. This was especially true of human-error pilots who took longer to answer questions. Their scores varied from a high of more than 86, for the median times, to less than 83 for those taking longer to answer the questions.

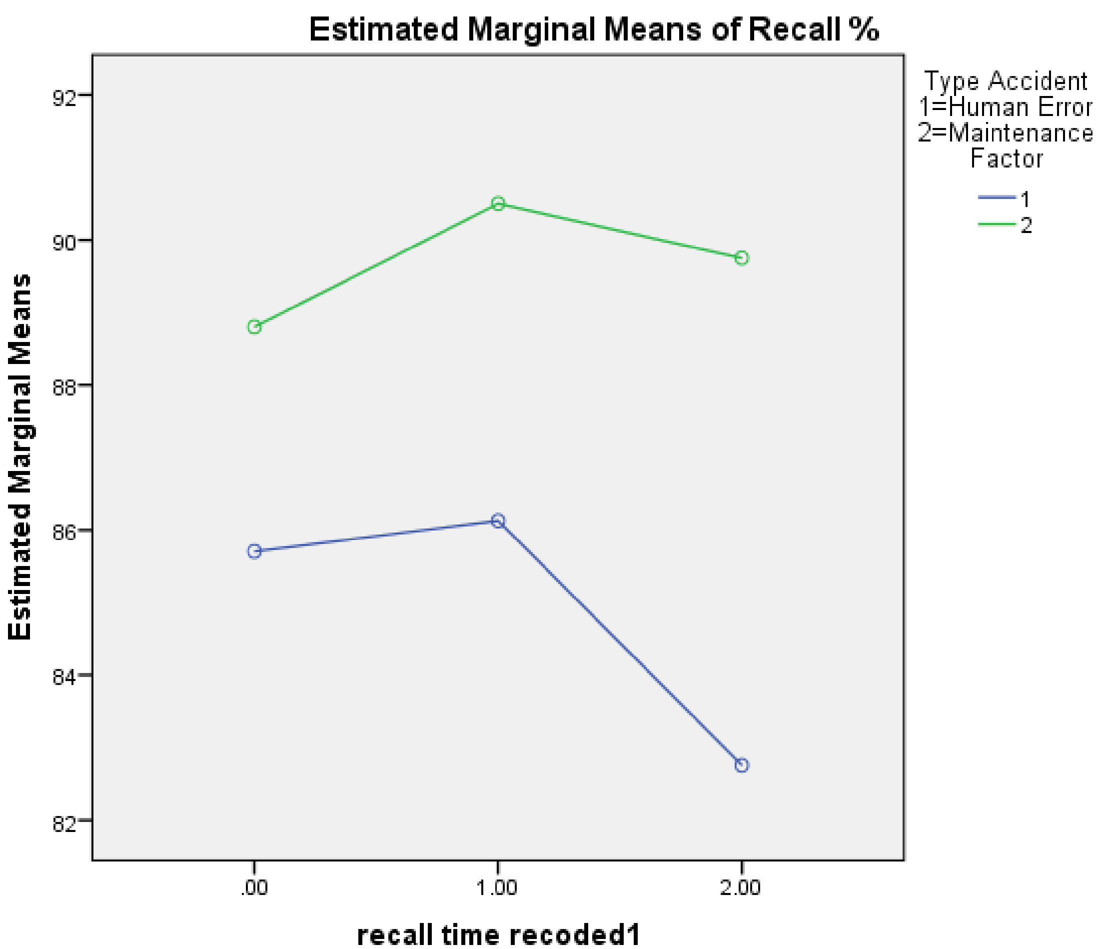

Figure 4.1 Estimated Marginal Means

In looking at Table 4.19, the F-Ratio is not significant, since the $\rho > .05$. This means that the time was similar. We conclude that, other things being equal, the human-error and the maintenance-factor group took approximately the same time to answer recall questions.

Table 4.17

Tests of Between-Subjects Effects Dependent Variable: Recall %

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	439.668[a]	5	87.934	1.037	.400
Intercept	264837.060	1	264837.060	3122.837	.000
TypeAccident	201.905	1	201.905	2.381	.126
recall1	21.843	2	10.921	.129	.879
TypeAccident * recall1	30.163	2	15.082	.178	.837
Error	9413.528	111	84.807		
Total	862504.000	117			
Corrected Total	9853.197	116			

In looking at the Helmert contrast on effects of time, we find that $\rho > .05$. The confidence level also crosses 0, which is another indication that it is not statistically significant.

Table 4.18

Contrast Results (K Matrix) Dependent Variable Recall %

	Helmert Contrast		
Level 1 vs. Level 2	Contrast Estimate		-4.819
	Hypothesized Value		0
	Difference (Estimate – Hypothesized)		-4.819
	Std. Error		3.123
	Sig.		.126
	95% Confidence	Lower Bound	-11.009
	Interval for Difference	Upper Bound	1.370

In looking at the post hoc tests, the Bonferroni and Ryan-Einot-Gabriel-Welsch Range results showed that the differences in time are not significant, in that the $\rho > .05$.

Table 4.19

Multiple Comparisons Dependent Variable: Recall %

	(I) recall time recoded1	(J) recall time recoded1	Mean Difference (I-J)	Std. Error	Sig.
Bonferroni	**.00**	**1.00**	**-.24**	**2.060**	**1.000**
		2.00	**2.59**	**2.113**	**.669**
	1.00	**.00**	**.24**	**2.060**	**1.000**
		2.00	**2.83**	**2.088**	**.535**
	2.00	**.00**	**-2.59**	**2.113**	**.669**
		1.00	**-2.83**	**2.088**	**.535**

Table 4.20

Recall %

	recall time recoded1	N	Subset 1
Ryan-Einot-Gabriel-Welsch Range[a]	**2.00**	**37**	**83.51**
	.00	**39**	**86.10**
	1.00	**41**	**86.34**
	Sig.		**.387**

Note. Means for groups in homogeneous subsets are displayed. Based on observed means. The error term is Mean Square(Error) = 84.807. a. Alpha = .05.

Research Question 4

Research question four looked to see if there is any relation between questions that required calculation to answer and the time taken to answer the questions. In looking at the descriptive statistics, the mean of the human-error group was 81.6%, while the mean of the maintenance-factor group was 81.3%. For those pilots involved in human-error accidents, those who took shorter times to answer questions attained a higher percentage of correct answers. For those involved in maintenance-factor accidents, the opposite trend was observed. For this group, the longer they took to answer the questions, the higher their scores. This result has to be tempered with the observation that there were only 11 accidents of this type; in dividing them by time, only one accident fell in the longest-time category. The standard deviation of the calculate percentage consistently fell between 12 and 14. For the Levene's test, $F(5, 110) = .518$, $\rho > .05$, which was not significant.

Table 4.21

Dependent Variable: Calculate %

Type Accident	calculate time recoded	Mean	Std. Deviation	N
1	.00	84.37	12.345	38
	1.00	82.03	12.131	31
	2.00	78.44	14.502	36
	Total	81.65	13.181	105
2	.00	76.67	13.429	3
	1.00	80.71	14.009	7
	2.00	100.00	.	1
	Total	81.36	13.981	11
Total	.00	83.80	12.414	41
	1.00	81.79	12.305	38
	2.00	79.03	14.732	37
	Total	81.62	13.196	116

Note. Human Error=1; Maintenance Factor=2.

Table 4.22

Levene's Test of Equality of Error Dependent Variable: Calculate %

F	df1	df2	Sig.
.850	5	110	.518

Note. Variances tests the null hypothesis that the error variance of the dependent variable is equal across groups.

To answer this research question, a factorial ANOVA was set up using the Calculate % as the dependent variable and the type of accident and the calculated time transformed into a nominal numeric input as the independent variables. The resulting test of between-subject effects shows ρ > .05. This difference was not significant.

Table 4.23

Tests of Between-Subjects Effects Dependent Variable: Calculate %

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	1072.516[a]	5	214.503	1.245	.293	.054	6.225	.428
Intercept	161423.510	1	161423.510	936.885	.000	.895	936.885	1.000
TypeAccident	100.571	1	100.571	.584	.446	.005	.584	.118
calculate1	236.226	2	118.113	.686	.506	.012	1.371	.163
TypeAccident calculate1	620.225	2	310.113	1.800	.170	.032	3.600	.369
Error	18952.794	110	172.298					
Total	792810.000	116						
Corrected Total	20025.310	115						

Note. TypeAccident; Human Error=1; Maintenance Factor=2; Not Determined=3.
a. R Squared = .054 (Adjusted R Squared = .011) b. Computed using alpha = .05

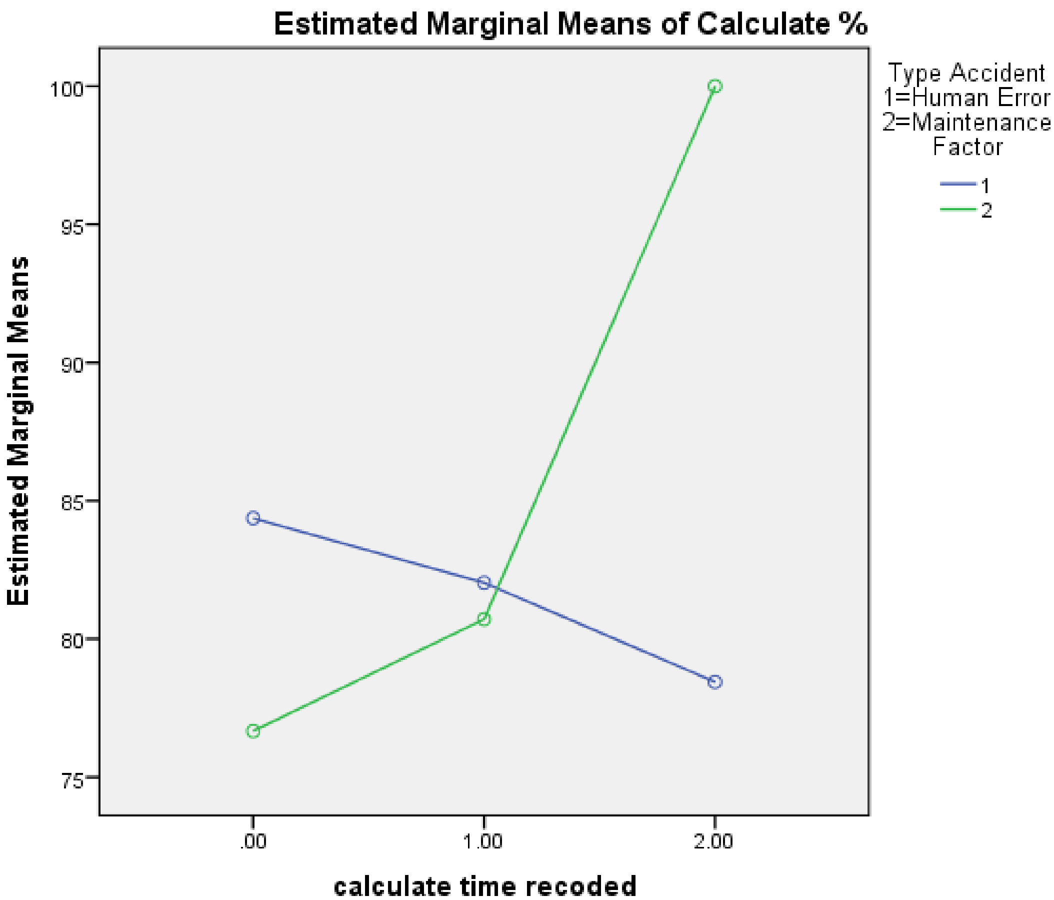

Figure 4.2 Estimated Marginal Means

Table 4.24

Contrast Results (K Matrix) Dependent Variable Calculate % Helmert Contrast

Level 1 vs. Level 2		
Contrast Estimate		-4.179
Hypothesized Value		0
Difference (Estimate – Hypothesized)		-4.179
Std. Error		5.469
Sig.		.446
95% Confidence Interval for Difference	Lower Bound	-15.018
	Upper Bound	6.660

Table 4.25

Calculate %

	calculate time recoded	N	Subset 1
Ryan-Einot-Gabriel-Welsch Range[a]	**2.00**	**37**	**79.03**
	1.00	**38**	**81.79**
	.00	**41**	**83.80**
	Sig.		**.265**

Note.Means for groups in homogeneous subsets are displayed. Based on observed means. The error term is Mean Square(Error) = 172.298.a. Alpha = .05.

Chapter 5 - Summary and Discussion of the Results

This final chapter is designed to summarize the design and methodology and to discuss the results with implications and possible future research recommendations. The study was conducted to see if there were any connectivity between the results of FAA knowledge tests and those GA pilots involved in fatal accidents. Four areas were investigated. The first was a comparison of knowledge-test results of those pilots involved in accidents (with pilot-error origins) to the test results of pilots with accidents involving maintenance factors or which were of undermined origin. The second research area involved the length of time per question that it took pilots to complete the knowledge test. The third area involved looking at the length of time it took pilots to answer recall questions and their relative success in doing so. The last research question looked at the time per question it took pilots to answer calculate questions and their success in answering them correctly.

In looking at the descriptive statistics, 4.3% of the 117 pilots involved in the study had failed a knowledge test between eight days and 30 months of their fatal accidents. All were involved in human-factors accidents, which results in 4.7 % of human-factors error pilots flying after failing a knowledge test. In her dissertation about accidents and fatalities in World War II, Pierce (2013) found evidence that the Aircrew Classification Battery, which was used to determine who became pilots, navigations, or bombardiers, also was a good indicator of pilot-error accidents. Those with the highest scores were less likely to be involved in this type of accident. Because these pilots demonstrated a lack of knowledge for the next rating, the FAA may want to institute a policy that requires a check of a pilot's knowledge for the rating that they hold if they fail a knowledge test. The FAA may consider additional flight training following failure of a knowledge test or even denial of flight privileges until the pilot can demonstrate proficiency in an aircraft to an FAA inspector, Designated Pilot Examiner (DPE), FAASTeam member, or CFI designated by the FSDO.

In examining the types of questions answered correctly, pilots answered 85.3% of the recall questions but only 81.5% of the calculate questions. This could indicate that the pilots did not have an in-depth understanding of the material, so they were not able to calculate the correct answer. Another possibility was their methods of preparation for the test. If they engaged in memorization of the answers in the practice tests provided by commercial vendors, they would be in a position to answer the recall questions but may not have understood the material sufficiently to be able to answer the calculate questions. Previous studies have shown that information that is memorized in a rote manner cannot be remembered as well as material that has been studied and is more meaningful (Craik and Lockhart, 1972).

The FAA has traditionally used a passing score of 70% of the total questions asked on the test. A pilot could fail all the questions in a particular area and still pass the test. The FAA may want to consider raising the minimum passing score on an examination. An alternative may be available with the implementation of Airmen Certification Standards (ACS). It may be possible to combine ACS with adaptive testing. If a pilot chooses an incorrect answer on a knowledge question on a particular task, additional questions would be asked within that task. If, after additional questioning, the algorithm determined that the task was unsatisfactory, as with the current PTS, it would make the area of operation unsatisfactory, and the applicant would be given a failure for the test. As in current practical testing, the applicant could be notified that he has failed the test, and the test could be terminated at that point, or he could be allowed to continue to take the test, to see if he can pass the other areas of operation. Test results would be by area of operations rather than one score for the entire test. During retesting, it might be appropriate to test only the area of operation that the applicant failed. This would reduce the number of questions asked and reduce the exposure of the questions to applicants who may memorize the questions and pass them on to commercial test-preparation companies. This would make FAA testing similar to JAA testing which involves seven separate knowledge tests that have to be passed (Casner et al. 2004). In this case the applicant would have to pass each of the appropriate areas of operation which would be combined into one test.

There should be more of an attempt to preserve the integrity of the test. The NCSBN has much tighter controls over test information disclosure than does the FAA (AFS 630, 2012). The council copyrights the test and makes cheating a criminal offense. The maximum FAA penalty includes the possibility of revocation of ratings for one year for cheating. The definition of cheating might be expanded to include passing on questions asked during the test. While criminal penalties might be a possibility, if the test were copyrighted, this would probably be excessive.

Discussion of Research Findings

The results of the first research question initially seemed to be a significant finding, when the mean of the human-error knowledge test results was 84% that of the maintenance-related accidents was almost 87%. While continuing the analysis, the Levene's test $F(2, 130) = .120$, $\rho > .05$. This shows that, while there is a difference in the average knowledge test result, it is not significant. An attempt was made to look at the results of an individual type of test, but because there were 27 types of test involved in the results, the sample sizes were so small that meaningful data could not be obtained. The test results of pilots involved in maintenance accidents came from seven different types of tests with no more than two pilots involved in accidents from any specific type of test. The largest number of tests was the instrument rating airplane with 26 tests; of these, 24 were from pilots involved in human-error type accidents, and only two were from those involved in maintenance-error accidents. Because the time allowed and number of questions assigned to each type test varied, it was not possible to use the total times for the tests in any meaningful comparison in question number two. The average calculate time and recall times per question were used for this research question. Unexpectedly, the average time per question was more for pilots involved in human-error accidents than it was for those involved in maintenance-related accidents for both calculate and recall times. It had been assumed that the times for pilots in human-factors error would have been less. The results were not significant as Levene's test $F(2, 130) = .181$, $\rho > .05$. It was thought that they might be more superficial learners, using rote memorization for answers, and therefore would have answered more quickly. In research question number three, the recall correct percentage was used as the dependent variable; the type of accident and the calculated time transformed into a nominal numeric input were the fixed factors. The mean for the pilots involved in human-error accidents was 84.94, whereas the mean for those involved in maintenance errors was 89.45. For the Levene's test, $F(5, 111) = .003$ $\rho < .05$. When the average time to answer a recall question is ignored, those involved in maintenance-related accidents scored significantly better than those who were involved in human-error related accidents. It could be that the pilots in the maintenance-factor group were deep learners. They may have understood the material better and were able to apply their knowledge when confronted with questions that were similar but not the exact wording of those that they may have studied. In the fourth research question, the average calculate time was used as the dependent variable, and the type of accident and the average calculate time per question were the fixed factors. The mean for those involved in human-error accidents was 81.65, whereas the mean for those involved in maintenance accidents was 82.36. For the Levene's test, $F(5, 110) = .518$ $\rho > .05$. It can be concluded that, other things being equal, those involved in accidents of both human factors and maintenance-related errors took the same amount of time to answer calculate questions and received similar scores.

Implications for Safety

The GA fatality rate has been slowly improving as aircraft have become more reliable and TAA provide the pilot much more situational awareness. This has been somewhat offset by the fact that TAA aircraft are more complex and, by themselves, do not lower accident rates; they can cause information overload for some pilots. FITS training programs, when used in conjunction with TAA, appears to improve risk management and develop self-assessment skills, which lead to life-long learning and a focus on safe mission accomplishment rather than individual tasks. Wright (2010) believed that concentrating on risk management and other higher-order skills is necessary to improve safety. He faulted the FAA for concentrating on airline pilots when it is the GA curriculum that needs attention. These self-assessment skills are important, as many accidents result from pilots' overconfidence in their knowledge or their ability to handle situations that may arise. Rakovan et al. (1999) determined in a study that a group of pilots not involved in accidents flew with a certified flight instructor more often than a similar group of pilots who had been involved in an accident. Casner et al. (2006) found that being involved in a flight review or preparing for a practical test created the climate for better retention.

Implications for Testing

Semb and Ellis (1994) found that there is not an appreciable difference in the forgetting curve between higher- and lower-ability students and that you can increase retention by increasing the amount of original learning. Orasanu et al. (2001) found that relevant knowledge was essential to good decision making. Lack of current knowledge can lead to pilots underestimating risks. Wright (2010a) believed that concentrating on risk management and higher-order skills is the way to improve safety. It may be that testing should reduce the number of recall questions and add questions of the calculate or interpret type, which could include risk analysis and scenario-type questions that could more thoroughly test higher-order skills. Hunter (2003) believed that SJT might prove useful in the certification of pilots. Goyer (2010) believed that questions should relate to what a pilot is doing in the cockpit and eliminate many of the rote questions that do not improve safety but only show that the pilot had a good instructor helping to prepare for the test. If adding SJT and more calculate or interpret questions would cause tests to exceed the number of questions that are currently in each test, a possible solution would be to use adaptive testing. This technique can reduce the number of questions in a test (Thompson and Weiss, 2011), whereas some authors believe that it more accurately measures those test takers whose grade is farther from the mean.

Recommendations for Future Research

A study could help to see if a system of continuing education makes for safer pilots. There are two programs that involve continuing education. Both have previously been mentioned: the WINGS program and the master certified flight instructor program. With the cooperation of the FAA, a study could be devised so that a group of pilots participating in the

WINGS program is paired with a group of pilots who are not in the program, and their safety records could be compared over a period of time. Because we are dealing with a sufficiently large pilot population—about 600,000 (Federal Aviation Administration, 2014),—a random sample of those who are active in the WINGS program and those who are not could be taken and statistically compared, to see if there is any difference in the safety records of the two populations.

Master instructors go through a process of continuing education, service, media development, participation, and teaching. There have been more than 1,700 applications for this designation, but many of the masters have received the designation as many as nine times. A study involving the 158 current master instructors (Hill and Hill, 2014) who are listed on the Master Instructor LLC website and a group of instructors who have not gone through this process may prove valuable. Instructors can renew their instructor ratings by: passing a practical test, with a record of recommending at least five students for ratings and achieving a passing rate of 80%; having a record of being a company check pilot or some position requiring the regular evaluation of pilots; taking an approved flight instructor refresher course (FIRC); or being in the military and passing an armed forces military instructor proficiency check. Some of these methods of renewal require little or no continuous educational activity between checks. The FIRC is often used by instructors who are no longer active but do not want to allow their flight-instructor ratings to lapse. Each attends a 16-hour course and receives a certificate to submit with the application to renew the rating. They do not have to demonstrate proficiency in an aircraft. There is a testing component to the FIRC (Federal Aviation Administration, 2011). It consists of a minimum of 30 multiple-choice questions. At least 15 of the questions must be at the end of the training; the other 15 questions can be spread throughout the 16 hours of the program. If a pilot does not meet the minimum of 70%, he or she can retrain until reaching 100%. Only deficient areas need to be retested, but the same questions cannot be used.

There have been 1,700 applications for master educator since 1997 (Hill and Hill, 2014). There are currently about 98,000 certified flight instructors, according to the Federal Aviation Administration (2014). Each master instructor's record could be paired with an instructor of similar years of experience and flight time; those data could be analyzed to see if there are any significant differences in the accident history of the two groups.

Another possible study that could be of benefit is to look at the accident history of the pilots who have earned at least one phase of the WINGS program and compare them with a representative sample of those who have not participated in the WINGS program. The FAA estimates that there are 599,086 active pilots (Federal Aviation Administration, 2014), so there is a large sample from which to draw. Those who have participated in a phase of the WINGS program have demonstrated a level of concern for safety that may not be present in the 575,000 pilots who have chosen not to participate in the program.

Another area that could be examined is a longitudinal study involving retraining or recurrent training for a group of GA pilots, along the lines of that done by the airlines. This training could be done annually and involve fewer than the 25 hours required by the airlines. It might be done online or with local seminars using a training approach similar to the WINGS program, so that there would be few travel expenses for the academic portion of the training. Instead of expensive level C and D simulators, less expensive flight-simulator training devices (FSTDs), found at many fixed-base operators (FBOs) or collegiate aviator programs, could be used. While FSTDs are used primarily for instrument training, many of them have visual systems that could be used for private pilots without instrument ratings. More frequent flying with a flight instructor was a distinguishing factor found by Rakovan, et al. (1999) in a non-accident group, compared to a group that had been involved in accidents. This study would see if similar results could be attained.

The final proposal recommended for future study is CAT or adaptive testing. Using CAT appears to provide an economy of questions and time and an increase in validity and reliability. A study just comparing results of the current fixed form versus CAT would probably not yield any meaningful results. If the questions in a test using CAT were truly representative of cockpit activities, used SJT, and truly measured higher-order skills, it may be a predictive tool similar to the Aircrew Classification Battery that Pierce (2013) found as an accurate predictor of pilot accident rates during World War II.

References

Adams, R. J. (1992). *Introduction of cognitive processes of expert pilots.* (Final No. DOT/FAA/RD-92/12). Washington D.C.: Research and Development Service.

AFS 630. (2012). *NCSBN and AFS-630 telecon 11/13/2012.* Unpublished manuscript.

Amalberti, R. (1993). Safety in flight operations. *Reliability and Safety in Hazardous Work Systems; Approaches to Analysis and Design,* 171-194.

Amalberti, R., & Wibaux, F. (2000). Advanced automation glass cockpit certification. *Human Factors in Certification,* 361-374.

American Composites Manufacturers Association. (2015). Maintaining your CCT certification. Retrieved from http://www.acmanet.org/cct/recertification

Bashook, P. G., & Parboosingh, J. (1998). Recertification and the maintenance of competence. *British Medical Journal, 316*(7130), 545-548.

Beard, B. L., & Geven, C. T. (2005). *Enhanced general aviation decision making: Weather assessment during preflight planning.* http://www.hf.faa.gov/docs/508/docs/gaFY05EDM.pdf: Federal Aviation Administration.

Billing, D. (2007). Teaching for transfer of core/key skills in higher education: Cognitive skills. *Higher Education, 53*, 483-516.

Boedigheimer, D. (2010). *Exploring reactions to pilot reliability certification and changing attitudes on the reduction of errors.* (Doctor of Philosophy, Northcentral University). , 1-159.

California State University, Long Beach. (2014). Critical values of fmax for Hartley's homogeneity of variance test. Retrieved April 26, 2014, Retrieved from http://www.csulb.edu/~acarter3/course-biostats/tables/table-Fmax-values.pdf

Campbell, R. D., & Bagshaw, M. (2002). *Human performance and limitations in aviation* (Third ed.). Ames, IA: Blackwell Science Ltd.

Casner, S. M., Heraldez, D., & Jones, K. M. (2006). Retention of aeronautical knowledge. *International Journal of Applied Aviation Studies, 6*(1), 71-97.

Casner, S. M., Jones, K. M., Puentes, A., & Irani, H. (2004). *FAA pilot knowledge tests: Learning or rote memorization.* (Technical Memorandum No. NASA/TM-2004-212814). Moffett Field, CA: National Aeronautics and Space Administration.

Cole, W. L. (1967). *General aviation public education program.* (Final No. FAA-FS-67-1). New York, NY: Flight Safety Foundation, Inc.

Craig, P. A. (2013). In Hager L. S. (Ed.), *The killing zone: How and why pilots die* (2nd ed.). New York, NY: The McGraw-Hill Companies Inc.

Craik, F., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.

Croft, J. (2012). GA safety remains in NTSB crosshairs. *The Weekly of Business Aviation, 95*(20), 3.

de Jong, T., & Ferguson-Hessler, M. G. M. (1996). Types and qualities of knowledge. *Educational Psychologist, 31*(2), 105-113.

Dekker, S. W. A. (2001). Follow the procedure or survive. *Human Factors and Aerospace Safety, 1*(4), 381-385.

Delta Corporate Communications. (30 Sep 2013). Delta to equip 11,000 pilots with Microsoft Surface 2 tablet devices. Retrieved from http://ww1.jeppesen.com/company/newsroom/articles.jsp?newsURL=news/newsroom/2013/Delta_Surface2EFB_NR.jsp

Detwiler, C. A., Hackworth, C. A., Holcomb, K. A., Boquet, A. J., Pfleiderer, E., Wiegmann, D. A., & Shappell, S. A. (2006). *Beneath the tip of the iceberg: A human factors analysis of general aviation accidents in Alaska versus the rest of the United States.* (No. DOT/FAA/AM-06/7). Washington, DC: Office of Aerospace Medicine.

Driskill, W. E., Weissmuller, J. J., Quebe, J. C., Hand, D. K., & Hunter, D. R. (1998). *Evaluating the decision-making skills of general aviation pilots.* (Final No. DOT/FAA/AM-98/7). Washington, DC: Office of Aerospace Medicine.

Ellis, J. A., Semb, G. B., & Cole, B. (1998). Very long-term memory for information taught in schools. *Contemporary Educational Psychology, 23*(4), 419-433.

Farr, M. J. (1987). *The long-term retention of knowledge and skills: A cognitive and instructional prospective*. New York, NY: Springer-Verlag.

Fata-Hartley, C. (2011). Resisting rote: The importance of active learning for all course learning objectives. *Journal of College Science Teaching, 40*(3), 36-39.

Federal Aviation Administration. Airplane simulator qualification, Advisory Circular 120-40b (1991).

Federal Aviation Administration. (2009). *Pilot's handbook of aeronautical knowledge*. Washington, DC: U.S. Department of Transportation.

Federal Aviation Administraion. PRIVATE PILOT AIRPLANE airman certification standards (draft), FAA-S-8081-ACS-PA (TBD).

Federal Aviation Administration. (2011). Nationally scheduled FAA approved industry conducted flight instructor refresher course, AC 61-83G.

Federal Aviation Administration. (2011a) WINGS - pilot proficiency, AC 61-91J.

Federal Aviation Administration. (2012a). Answering the challenge to improve aviation safety. Retrieved December 16, 2012, Retrieved from http://www.faa.gov/about/office_org/headquarters_offices/avs/offices/afs/afs600/media/reg_support_brochure.pdf

Federal Aviation Administration. (2012b). *Aviation rulemaking advisory committee (ARAC); new task assignment for the ARAC: establishment of airman testing standards and training working group.* (Reports No. 1039101936). Lanham: Federal Information & News Dispatch, Inc.

Federal Aviation Administration. (Page last modified: July 10, 2014 4:53:27 PM EDT). US civil airmen statistics. Retrieved from http://www.faa.gov/data_research/aviation_data_statistics/civil_airmen_statistics/2013/

Federal Aviation Administration. Certification: Pilots, Flight Instructors, and Ground Instructors Title 14 U.S.C. Part 61 (2014a).

Federal Aviation Administration. Operating requirements: Domestic, flag, and supplemental operations, Title 14 U.S.C. Part 121 (2014b).

Federal Aviation Administration. (2014c). Safer skies through education. Retrieved from http://www.faasafety.gov/default.aspx

Federal Aviation Administration. (2014d). Applicant identification, information verification, & authorization requirements matrix. Retrieved from http://www.faa.gov/training_testing/testing/media/testing_matrix.pdf

Federal Aviation Administration. (Downloaded January 4, 2015). Eastern Airlines flight 401, Lockheed model L-1011, N310EA. Retrieved from http://lessonslearned.faa.gov/ll_main.cfm?TabID=3&CategoryID=9&LLID=8

Field, A. (2009). *Discovering statistics using SPSS* (Third Edition ed.). London, England: SAGE Publications Ltd.

Frantz, R. (2011). The need for recertification. *IEEE Communications Magazine, 49*(3), 12.

Goh, J., & Wiegmann, D. A. (2002). Human factors analysis of accidents involving visual flight rules into adverse conditions. *Aviation Space and Environmental Medicine, 73*(8), 817-822.

Goldman, S. M., Fiedler, E. R., & King, R. E. (2002). *General aviation maintenance-related accidents: A review of ten years of NTSB data.* (Final No. DOT/FAA/AM-02/23). Washington, DC: Office of Aerospace Medicine.

Government Accountability Office. (2011). *Initial pilot training Better management controls are needed to improve FAA oversight.* (Report to Congressional Requestors No. GAO-12-117). Washington, DC: U.S. Government Accountability Office.

Goyer, R. (May 26, 2010). Why the FAA written tests suck and what you can do about it. Retrieved from http://www.flyingmag.com/blogs/going-direct/why-faa-written-tests-suck-and-what-you-can-do-about-it

Hackworth, C. A., King, S. J., Cruz, C., Thomas, S., Roberts, C., Bates, C., & Moore, R. (2007). *The private pilot practical test: Survey results from designated pilot examiners and newly certificated private pilots.* (Final No. DOT/FAA/AM-07/17). Oklahoma City, OK: FAA Civil Aerospace Medical Institute,.

Halleran, M. S., & Wiggins, M. E. (2010). Changing general aviation flight training by implementing FAA industry training standards. *The International Journal of Applied Aviation Studies, 10*(1), 117-130.

Harris, D., & Muir, H. C. (Eds.). (2005). *Contemporary issues in human factors and aviation safety*. Burlington, VT: Ashgate Publishing Company.

Helmreich, R. L., & Merritt, A. C. (1998). *Culture at work in aviation and medicine: National, organizational and professional influences*. Aldershot, England: Ashgate Publishing Limited.

Hennig, J. C. (2012). *Recommendations to enhance the airman knowledge test content and its processes and methodologies for training and testing.* (Airman Testing Standards and Training Aviation Rulemaking Committee). Washington, DC: Federal Aviation Administration.

Hise, P. (2002). *Pilot error: The anatomy of a plane crash* (First ed.). Dulles, VA: Brassey's, Inc.

Hopkins, J. (1994). Critical success factors. *Flying, 121*(10), 84.

Hubbard, T. P. (2000). *Effects of an inhibition Testing Model on private pilot, airplane single engine land, ground school students.* (Doctor of Education, Oklahoma State University), 1-262.

Hunter, D. R. (1995). *Airman research questionnaire: Methodology and overall results.* (No. DOT/FAA/AM-95/27). Washington, DC: Office of Aviation Medicine.

Hunter, D. R. (2003). Measuring general aviation pilot judgment using a situational judgment technique. *The International Journal of Aviation Psychology, 13*(4), 373-386.

Jensen, R. S. (1995). *Pilot judgment* (Third ed.). Burlington, VT: Ashgate Publishing Company.

Jensen, R. S., Guilkey, J. E., & Hunter, D. R. (1998). *An evaluation of pilot acceptance of personal minimums training program for risk assessment.* (Final No. DOT/FAA/AM-98/6). Washington, DC: Office of Aviation Medicine.

Kinaszczuk, T., Jr., Dodge, R. E., & Mohler, S. R. (1982). Safety attitudes of a general aviation population. *Aviation Space and Environmental Medicine, 53*(12), 1227-1229.

Knecht, W. (2010). Effects of video weather products on general aviation pilot weather knowledge and flight behavior into adverse weather. *The International Journal of Applied Aviation Studies, 10*(1), 169-202.

Knecht, W. R., & Lenz, M. (2010). *Causes of general aviation weather-related non-fatal incidents: Analysis using NASA aviation safety reporting system data.* (Final No. DOT/FAA/AM-10/13). Oklahoma City, OK: FAA Civil Aerospace Medical Institute.

Knecht, W., Harris, H., & Shappell, S. A. (2005). *The influence of visibility, cloud ceiling, financial incentive, and personality factors on general aviation pilots' willingness to take off into marginal weather, part I: the data and preliminary conclusions.* (Final No. DOT/FAA/AM-05/7). Washington, DC: Office of Aerospace Medicine.

Knecht, W. R., & Smith, J. (2012). *Effects of training school type and examiner type on general aviation flight safety.* (Final No. DOT/FAA/AM-13/4). Washington, DC: Federal Aviation Administration.

Konoske, P. J., & Ellis, J. A. (1991). Cognitive factors in learning and retention of cognitive tasks. *Instruction: Theoretical and Applied Perspectives,* 47-70.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessment. *Journal of Personality and Social Psychology, 77*(6), 1121-1134.

Lenné, M. G., Ashby, K., & Fitzharris, M. (2008). Analysis of general aviation crashes in Australia using the human factors analysis and classification system. *The International Journal of Aviation Psychology, 18*(4), 340-352.

Li, G. (1994). Pilot-related factors in aircraft crashes: A review of epidemiologic studies. *Aviation Space and Environmental Medicine, 64*(10), 944-952.

Lythgoe, T. J. (2011). Flight simulation for the brain: Why army officers must write. *Military Review, 91*(6), 49-57.

Maurino, D. E. (1999). Safety prejudices, training practices, and CRM: A midpoint perspective. *The International Journal of Aviation Psychology, 9*(4), 413-422.

McElhatton, J., & Drew, C. (1993). Time pressure as a causal factor in aviation safety incidents: The "hurry up" syndrome. *Proceedings of the Seventh International Symposium on Aviation Psychology; 22-20 April 1993,* Columbus, Ohio. 269-274.

Moebus, K., Moebus, P., Schwaninger, A., Koller, S., Iglesias, S., & Wales, A. (2009). *Impact assessment of the publication of questions of theoretical examinations for part 66 and part FCL.* (Final No. EASA.2008.C52). Zurich Airport, Switzerland: MOEBUS Aviation Consulting.

Moore, P. J., & Telfer, R. A. (1990). Approaches to learning: Relationships with pilot performance. *The Journal of Aviation/Aerospace Education and Research, 1*(1), 44-58.

Murray, S. R. (1997). Deliberate decision making by aircraft pilots: A simple reminder to avoid decision making under panic. *The International Journal of Aviation Psychology, 7*(1), 83-100.

National Transportation Safety Board. (Nov 14, 2012). Improve general aviation safety. Retrieved December 16, 2012. Retrieved from http://www.ntsb.gov/safety/mwl.html

National Transportation Safety Board. (2014). Accident database & synopses. Retrieved from http://www.ntsb.gov/aviationquery/index.aspx

Oakes, D. W., Ferris, G. R., Martocchio, J. J., Buckley, M. R., & Broach, D. (2001). Cognitive ability and personality predictors of training program skill acquisition and job performance. *Journal of Business and Psychology, 15*(4), 523-548.

O'Hare, D., & Chalmers, D. (1999). The incidence of incidents: A nationwide study of flight experience and exposure to accidents and incidents. *The International Journal of Aviation Psychology, 9*(1), 1-18.

O'Hare, D., & Douglas, O. (2002). Cross-country VFR crashes: Pilots and contextual factors. *Aviation Space and Environmental Medicine, 73*(4), 363-366.

O'Hare, D., Wiggings, M., Batt, R., & Morrison, D. (1994). Cognitive failure analysis for aircraft accident investigation. *Ergonomics, 37*(11), 1855-1869.

Orasanu, J., Martin, L., & Davison, J. (2001). Cognitive and contextual factors in aviation accidents: Decision errors. *Linking Expertise and Naturalistic Decision Making*, 209-225.

Ortner, T. M., & Caspers, J. (2011). Consequences of test anxiety on adaptive versus fixed item testing. *European Journal of Psychological Assessment, 27*(3), 157-163.

Patankar, M. S., Brown, J. P., & Treadwell, M. D. (2005). *Safety ethics cases from aviation, healthcare and occupational and environmental health*. Burlington, VT: Ashgate Publishing Company.

Pierce, M. R. (2013). *Earning their wings: Accidents and fatalities in the United States Army Air Forces during flight training in World War Two* (Unpublished Ph. D.). Kansas State University, Manhattan, KS.

Rakovan, L., Wiggings, M. W., Jensen, R. S., & Hunter, D. R. (1999). *A survey of pilots and the dissemination of safety information.* (Final No. DOT/FAA/AM99/7). Washington, DC: FAA.

Rolfe, J. M. (1972). Ergonomics and air safety. *Applied Ergonomics, 3*(2), 75-81.

Saleem, J. J. (1996). A case-based review of critical incidents in general aviation for improved safety. *The International Journal of Applied Aviation Studies, 6*(2), 271-282.

Sarter, N. B., & Alexander, H. M. (2000). Error types and related error detection mechanisms in the aviation domain: An analysis of aviation safety reporting system incident reports. *The International Journal of Aviation Psychology, 10*(2), 189-206.

Schvaneveldt, R., Beringer, D. B., Lamonica, J., Tucker, R., & Nance, C. (2000). *Priorities, organization, and sources of information accessed by pilots in various phases of flight.* (No. DOT/FAA/AM-00/26). Washington D.C.: Office of Aviation Medicine.

Semb, G. B., & Ellis, J. A. (1994). Knowledge taught in school: What is remembered? *Review of Educational Research, 64*(2), 253-286.

Shappell, S. A., & Wiegmann, D. A. (2003). *A human error analysis of general aviation controlled flight into terrain accidents occurring between 1990 and 1998.* (Final No. DOT/FAA/AM-03/4). Washington, DC: Office of Aerospace Medicine.

Smith, E. M. B. (1966). Pilot error and aircraft accidents. *Zentralblatt Fur Verkehrs-Medizin, Verkehrs-Psychologie, Lurt-Und Raumfahrt-Medizin, 12*, 1-13.

Smithsonian National Postal Museum. (2004). Fad to fundamental: Airmail in America. Retrieved November 28, 2013, Retrieved from http://www.postalmuseum.si.edu/airmail/pilot/pilot_old/pilot_old.html

Snyder, K. T., Sokoloff, T., & Bearden, M. (2003). The cockpit associate: A "Co-pilot in a box" for general aviation. *Digital Avionics Systems Conference, 2*, 12.C.3 - 12.

Stefansson, G., & Sigurdardottir, A. J. (2011). Web-assisted education: From evaluation to learning. *Journal of Instructional Psychology, 38*(1), 47-60.

Stewart, J. E. (2008). Locus of control as mediators of hazardous attitudes among aviations: A review and suggested applications. *The International Journal of Applied Aviation Studies, 8*(2), 263-280.

Stowell, R. (2008). The problem with flight training. *Aviation Safety,* (March), 4-8.

Taneja, N., & Wiegmann, D. A. (2001). Analysis of mid-air collisions in civil aviation. *45th Annual Meeting of the Human Factors and Ergonomics Society,* Santa Monica, CA. 1-5.

Thompson, Nathan A., & Weiss, David A. (2011). A Framework for the Development of Computerized Adaptive Tests. *Practical Assessment, Research & Evaluation*, 16(1). Available online: http://pareonline.net/getvn.asp?v=16&n=1.

Trollip, S. R., & Jensen, R. S. (1991). *Human factors for general aviation*. Engelwood, CO: Jeppesen Sanderson.

Vail, G. J., & Ekman, L. G. (1986). Pilot-error accidents: Male vs female. *Applied Ergonomics, 17*(4), 297-303.

Weiss, D. J., & Betz, N. E. (1973). *Ability measurement: Conventional or adaptive?* (Technical Report No. RR-73-1). Washington, DC: Division Office of Naval Research.

Wiegmann, D. A., & Shappell, S. A. (1997). Human factors analysis of post accident data: Applying theoretical taxonomies of human error. *The International Journal of Aviation Psychology, 7*(1), 67-81.

Woods, D. D., Johannesen, L. J., Cook, R. I., & Sarter, N. B. (1884). *Behind human error: Cognitive systems, computers, and hindsight.* (State of the art report No. CSERIAC SOAR 94-01). Wright Patterson AFB: CSERIAC Program Office.

Wright, R. A. (2010a). TAA training. *Aviation Safety,* (July), 12-15.

Wright, R. A. (2010b). Training for risk management. *Aviation Safety,* (November), 8-11.

Wright, R. A. (2011). Will training reform help to reduce fatals? *Aviation Safety,* (July), 8-12.

Wright, R. A. (2012). Ineffective practical test standards. *Aviation Safety,* (November), 4-8.