

# EMBRY-RIDDLE

## Aeronautical University™

### SCHOLARLY COMMONS

---

Dissertations and Theses

---

6-2012

## Analysis of the Impact of Scenario-Based Training on the Aeronautical Decision Making of Collegiate Flight Students

Mariko Genevieve Doskow

*Embry-Riddle Aeronautical University - Daytona Beach*

Follow this and additional works at: <https://commons.erau.edu/edt>



Part of the [Aviation Safety and Security Commons](#)

---

### Scholarly Commons Citation

Doskow, Mariko Genevieve, "Analysis of the Impact of Scenario-Based Training on the Aeronautical Decision Making of Collegiate Flight Students" (2012). *Dissertations and Theses*. 48.

<https://commons.erau.edu/edt/48>

This Thesis - Open Access is brought to you for free and open access by Scholarly Commons. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of Scholarly Commons. For more information, please contact [commons@erau.edu](mailto:commons@erau.edu).

ANALYSIS OF THE IMPACT OF SCENARIO-BASED TRAINING ON THE  
AERONAUTICAL DECISION MAKING OF COLLEGIATE FLIGHT STUDENTS

by

Mariko Genevieve Doskow

A Thesis Submitted to the College of Aviation Department of Applied Aviation Sciences  
in Partial Fulfillment of the Requirements for the Degree of  
Master of Science in Aeronautics

Embry-Riddle Aeronautical University  
Daytona Beach, Florida  
June 2012

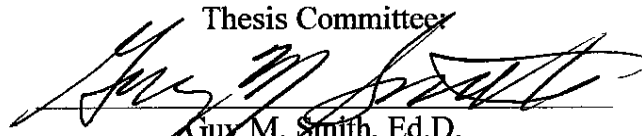
ANALYSIS OF THE IMPACT OF SCENARIO-BASED TRAINING ON THE  
AERONAUTICAL DECISION MAKING OF COLLEGIATE FLIGHT STUDENTS

by


Mariko Genevieve Doskow


This Thesis was prepared under the direction of the candidate's Thesis Committee Chair, Dr. Guy M. Smith, Associate Professor, Daytona Beach Campus, and Thesis Committee Members Michele Halleran, Associate Professor, Daytona Beach Campus, and Dr. Michael Wiggins, Professor, Daytona Beach Campus, and has been approved by the Thesis Committee. It was submitted to the Department of Applied Aviation Sciences in partial fulfillment of the requirements for the degree of Master of Science in Aeronautics

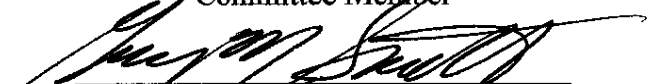
Thesis Committee


  
Guy M. Smith, Ed.D.  
Committee Chair

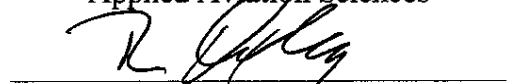
  
Michele S. Halleran, MSA  
Committee Member

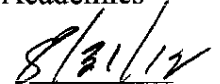
  
Michael Wiggins, Ed.D.  
Committee Member

  
John M. Lanicci, Ph.D.  
Graduate Program Chair  
Applied Aviation Sciences

  
Guy M. Smith, Ed.D.  
Department Chair  
Applied Aviation Sciences

  
Tim Brady, Ph.D.  
Dean, College of Aviation

  
Robert Oxley, Ph.D.  
Associate Vice President of Academics

  
Date

## Acknowledgements

This study would have not been possible without the support of several people. I would like to thank the thesis committee for their encouragement, advice, and judicious applications of tough love. Dr. Guy Smith, Prof. Michele Halleran, and Dr. Michael Wiggins – I could not have done this without you. I would also like to thank Dr. MaryJo Smith for her encouragement and statistical expertise, and Bruno Villela, for being a mentor and friend.

The support of family, friends, and loved ones was also essential to the success of this study. Thank you for listening to me, reminding me how badly I wanted this, and for not letting me give up.

Finally, I would like to thank my participants. Without your contributions, there would have been nothing to write about. I sincerely enjoyed working with you and watching you learn.

## Abstract

Researcher: Mariko Genevieve Doskow

Title: Analysis of the Impact of Scenario-Based Training on the Aeronautical Decision Making of Collegiate Flight Students

Institution: Embry-Riddle Aeronautical University

Degree: Master of Science in Aeronautics

Year: 2012

The persistence of faulty decision making as a primary cause of accidents indicates a need to train pilots to make better decisions. The purpose of this study was to analyze scenario-based training's effectiveness at improving the aeronautical decision making of collegiate flight students. The researcher scored each participant's aeronautical decision making as they completed simulated flights in an advanced aviation training device. The scores quantified the participants' aeronautical decision making on seven decision-making variables and served as the basis for generating an overall decision making score for each participant. The experimental group completed a scenario-based aeronautical decision making treatment between their simulated flights. Chronbach's alpha analyses verified the scoring's internal reliability. Mann-Whitney and Wilcoxon tests compared the participants' decision making before and after the experimental treatment. Although there were practical improvements, the differences were not statistically significant. The practical significance of the results suggests that further research is required.

## Table of Contents

	Page
Thesis Review Committee .....	ii
Acknowledgements.....	iii
Abstract .....	iv
List of Tables .....	x
List of Figures .....	xi
Chapter	
I    Introduction.....	1
Significance of the Study .....	4
Statement of the Problem.....	5
Purpose Statement.....	5
Hypothesis.....	5
Delimitations.....	5
Limitations and Assumptions .....	6
Definition of Terms.....	7
List of Acronyms .....	12
II    Review of the Relevant Literature .....	16
Judgment, ADM, and HOTS .....	16
SA (Formerly Situational Awareness).....	18
GA Pilots.....	19
ERAU Student Pilots .....	21
Expert ADM .....	21

	The Novice Pilot .....	21
	The Accident-Prone Pilot.....	22
	ADM Training .....	24
	Traditional ADM Training: Informal By-Product of Experience .....	24
	ADM: Behavior Management.....	25
	ADM: Problem-Solving Skills and Practice .....	29
	Problem-Based Learning .....	32
	Commercial and Military Training Implementation.. .....	32
	GA Training Research .....	35
	ADM Mnemonics and Acronyms .....	37
	Implementing GA ADM Training .....	37
	FAA CGAR .....	38
	FITS .....	39
	Continuing the GA Training Reform Initiative .....	43
	Summary .....	45
III	Methodology .....	47
	Research Approach .....	47
	Design and Procedures.....	47
	Designating the Control and Experimental Groups ... .....	47
	AATD Session Design .....	48
	AATD Session Procedure .....	49
	Treatment Design.....	50

	Treatment Procedure .....	50
	Apparatus and Materials .....	52
	Population Sample .....	53
	Data Collection Device .....	54
	Content Validity and Reliability .....	55
	Treatment of the Data .....	56
	Descriptive Statistics.....	56
	Reliability Testing.....	56
	Hypothesis Testing.....	57
IV	Results.....	58
	Descriptive Statistics.....	58
	Dependent Variables .....	60
	Reliability Testing.....	66
	Hypothesis Testing.....	67
	Baseline ADM Performance .....	67
	Ending ADM Performance .....	70
	Change in ADM for the Control Group.....	72
	Change in ADM for the Experimental Group.....	75
	Comparing Delta between Control and Experimental Groups.....	77
V	Discussion, Conclusions, and Recommendations.....	81
	Discussion .....	81
	Significance of Results .....	81
	Experiment Design Factors.....	81



Impact of Variations in the Sample .....	81
G1000 Proficiency .....	87
Types of ADM Tested. ....	87
In-flight ADM.....	87
Quick Decisions .....	88
Decision Points in Each Scenario .....	88
Unintended Decisions .....	91
Maintaining Realism of the Scenario.....	91
Deviations from the Scenario Procedure. ....	92
AATD-Induced Deviations from the Scenario Procedure. ....	92
Un-programmed engine Failure .....	92
Clouds and Visibility .....	93
Auto-Zoom.....	93
Operator-Induced Deviations from the Scenario Procedure .....	93
Treatment Design Factors .....	94
Researcher and Instrument Factors .....	95
Bias .....	95
Scoring .....	95
Instrument Reliability .....	96
Impact of Video Quality on Data Collection. ....	97
Conclusions.....	97
Recommendations.....	100

	Further Research and Analysis .....	100
	Suggested Improvements for this Study .....	100
	Experiment Design Factors.....	101
	Treatment Design Factors.....	101
	Bias .....	101
	References.....	102
Appendices		
A	Permission to Conduct Research .....	110
B	Participant Consent Form .....	112
C	Pilot Briefing.....	114
D	Sample Scoring Sheet .....	116
E	Sample Debrief Form.....	118
F	Sample AATD Session Procedure .....	120

## List of Tables

	Page
Table	
1 Ordinal Variables for the Control and Experimental Groups, First AATD Sessions .....	64
2 Ordinal Variables for the Control and Experimental Groups, Second AATD Sessions.....	65
3 Delta of Ordinal Variables for the Control and Experimental Groups .....	66
4 Overall ADM and Delta Overall ADM for Participants Grouped By Highest Certification: Control, Experimental, and All.....	82
5 Overall ADM and Delta Overall ADM for CFI and non-CFI Participants: Control, Experimental, and All .....	84
6 Overall ADM Scores and Delta Overall ADM, by Scenario Flown in Rounds One and Two .....	88
7 Practical Comparison of the Control and Experimental Groups, Overall ADM ...	99

## List of Figures

	Page
Figure	
1 Frasca Mentor Advanced Aviation Training Device (AATD) .....	53
2 Total Time of All Participants. ....	58
3 Total Time of the Participants, Control and Experimental Groups .....	59
4 Highest Certification Held by the Participants .....	60
5 Highest Certification Held by the Participants, Control and Experimental Groups .. .....	60
6 Participants Detected the Problem, Percentage of Decisions .....	62
7 Participants Resolved the Problem, Percentage of Decisions.....	62
8 Participants Reached a Safe Outcome .....	63
9 Round One Problem Comprehended Scores, Control and Experimental Groups .	68
10 Round One Problem Projected Scores, Control and Experimental Groups.....	68
11 Round One Decision Process Used Scores, Control and Experimental Groups ...	69
12 Round One Timely Manner Scores, Control and Experimental Groups .....	69
13 Round One Overall ADM Scores, Control and Experimental Groups.....	70
14 Round Two Problem Comprehended Scores, Control and Experimental Groups	70
15 Round Two Problem Projected Scores, Control and Experimental Groups.....	71
16 Round Two Decision Process Used Scores, Control and Experimental Groups ...	71
17 Round Two Timely Manner Scores, Control and Experimental Groups.....	72
18 Round Two Overall ADM Used Scores, Control and Experimental Groups .....	72
19 Problem Comprehended Scores in Round One and Round Two for the Control Group .....	73

20	Problem Projected Scores in Round One and Round Two for the Control Group .	73
21	Decision Process Used Scores in Round One and Round Two for the Control Group .....	74
22	Timely Manner Scores in Round One and Round Two for the Control Group.....	74
23	Overall ADM Scores in Round One and Round Two for the Control Group .....	75
24	Problem Comprehended Scores in Round One and Round Two for the Experimental Group.....	75
25	Problem Projected Scores in Round One and Round Two for the Experimental Group .....	76
26	Decision Process Used Scores in Round One and Round Two for the Experimental Group.....	76
27	Timely Manner Scores in Round One and Round Two for the Experimental Group .....	77
28	Overall ADM Scores in Round One and Round Two for the Experimental Group ..	77
29	Delta Problem Comprehended for the Control and Experimental Groups .....	78
30	Delta Problem Projected for the Control and Experimental Groups .....	78
31	Delta Decision Process Used for the Control and Experimental Groups .....	79
32	Delta Timely Manner for the Control and Experimental Groups .....	79
33	Delta Overall ADM for the Control and Experimental Groups.....	80
34	Total Time and Overall ADM for All Participants in Round One .....	85
35	Total Time and Overall ADM for the Control Participants in Round Two.....	85
36	Total time and Overall ADM for the Experimental Participants in Round Two...	86

## Chapter I

### Introduction

Human error continues to be a leading cause of General Aviation (GA) accidents and incidents. The Federal Aviation Administration (FAA, 1991) attributed 52% of fatal GA accidents to pilot error. The National Transportation Safety Board (NTSB) cited personnel-related causes or factors in 91% of GA accidents in 2006 (NTSB, 2006). The *2010 Nall Report* cited 70% of non-commercial fixed-wing accidents in 2009 and 63% of fatal non-commercial fixed-wing accidents in 2009 as pilot-related (Aircraft Owners and Pilots Association [AOPA], 2010). The 2009 rate of pilot-related accidents – 4.63 per 100,000 flight hours – was consistent with the rate of pilot-related accidents in 2008 and for the period, 2000-2008 (AOPA, 2009; AOPA, 2010).

The NTSB subdivided personnel-related causes of accidents into human performance issues such as aircraft control and handling, planning and decision-making, and use of aircraft equipment (NTSB, 2006). Of the accidents in 2006 for which the NTSB cited a human performance cause or factor, “the most frequently cited cause/factor was aircraft handling and control (71%), followed by planning and decision-making (36%)” (NTSB, 2006, p. 48). The *2010 Nall Report* divided pilot-related accidents into different categories than the NTSB. There were no categories related specifically to decision-making or judgment, but AOPA discussed decision-making’s impact on many of the categories it used to describe GA accidents. The report stated that “the judgment leading to any pilot-related accident could be called into question” (AOPA, 2010, p. 17). Fuel-management and weather accidents were singled out as being “primarily [...] failures of flight planning and in-flight decision-making” (AOPA, 2010, p. 17). The

report mentioned the possibility of pilots underestimating the risks associated with the takeoff phase of flight as a contributing factor to the high number of takeoff phase accidents (AOPA, 2010). Many maneuvering accidents also resulted from risky maneuvers initiated at low altitudes. The majority began with a loss of control or stall at altitudes too low to recover, indicating that “these accidents were more tied to poor judgment than lack of knowledge or skill” (AOPA, 2010, p. 24).

Decision-making errors may be under-reported, even when they are identified as a separate category. The *2010 Nall Report* did not provide statistics for how many accidents resulted from poor risk management or faulty aeronautical decision making (ADM) (AOPA, 2010). The NTSB’s reviews reported that 36% of personnel-related GA accidents were caused by poor planning or decision-making, but coded each accident with a single defining event code instead of performing root cause analysis and reporting each of the causes and factors found (NTSB, 2006). Meanwhile, a recent study (Wright, 2009) applied root cause analysis to 29 fatal accidents involving a popular GA aircraft. The study concluded that 25 of those accidents could have been avoided using fundamental risk management procedures or higher order thinking skills (HOTS) such as ADM and single pilot resource management (Wright, 2009). Only four of those accidents resulted from the pilot’s faulty aircraft handling (Wright, 2009).

The search for a better method of acquiring judgment in aviation led to a significant body of research on judgment or ADM and how to train it. Since Jensen’s 1982 study, the premise that judgment can be taught has been accepted in academia and commercial aviation (FAA, 1991). GA, however, has been slow to accept that judgment can be taught or come to a consensus on how to train ADM. The FAA has also been slow

to provide guidance on how to provide ADM training. Much of the guidance the FAA provided had not been updated as recently as the early 2000s (“FAA-Industry,” 2003).

Significant efforts to improve formal ADM training in GA include projects by the FAA Center For General Aviation Research (CGAR), the FAA Industry Training Standard (FITS) program, and the Society of Aviation and Flight Educators (SAFE). CGAR is a consortium of aviation universities conducting research to make significant improvements in safety and efficiency for GA air transportation (CGAR, 2005). The FITS program is a collaboration of FAA, industry, and the FAA Center of Excellence for GA. FITS formed with a mission to improve safety by reducing human error in GA with a new training philosophy that accelerates the acquisition of higher-level judgment and decision-making skills (FITS Master Instructor Syllabus, 2006). Goals include developing adaptive training and industry standards for the GA community (FAA-Industry Training Standards [FITS] program plan, 2003). SAFE is an organization of aviation educators “fostering professionalism and excellence in aviation through continuing education, professional standards, and accreditation” (“About SAFE,” 2012, para. 1). SAFE’s Mission Statement states that they seek to “create a safer environment through enhanced education” (“Vision & Mission Statement,” 2012, Mission Statement, para. 1).

New training strategies that emphasize ADM and other mental skills in GA training have been proposed as the key to meaningfully reforming the entire GA training paradigm (SAFE, 2011a). GA flight training remained mostly unchanged from the maneuvers-based focus of the Civilian Pilot Training (CPT) program as recently as 2009.



Meanwhile GA accident rates – particularly for decision-making, human error accidents – remained stagnant, illustrating the need to reform GA training (Wright 2009).

FITS researchers argued that implementing a scenario-based training (SBT) paradigm in GA training was imperative to better prepare GA pilots because the flight environment is becoming increasingly challenging as well (“FITS Master,” 2006). Higher performance technologically advanced aircraft (TAA) are increasingly putting pilots with less experience and training into situations that require flight management and decision-making skills normally expected from air transport pilot (ATP) certificated pilots (“FITS Master,” 2006). Evolution of technology in GA aircraft such as displays and automation has rapidly outpaced training programs and the guidance, standardization, and certification (GSC) provided by the FAA. This increasing disparity exacerbates the current GA training paradigm’s deficiency in teaching adequate ADM (“FITS Master,” 2006).

### **Significance of the Study**

Improving the ADM of GA pilots would have a significant impact on the safety of individual GA pilots and on the health of the greater GA community. Faulty ADM contributed to a significant percentage of past fatal GA accidents; improving ADM ability in GA pilots could prevent many future fatal GA accidents (“FAA-Industry Training Standards (FITS) Program Plan,” 2003; Wright, 2009). Improved GA safety would have the additional benefit of improving the general public’s perception of GA safety which would enable GA growth (Wright, 2009).

## **Statement of the Problem**

Human errors in judgment continue to be a leading cause of aviation accidents and incidents while GA accident rates have failed to improve significantly during the last decade (AOPA, 2010). Stagnant accident rates indicate that the current training system in GA has reached the limits of its usefulness for training safer pilots. The persistence of faulty decision making as a primary cause of human error and pilot-related accidents indicates a specific need to train pilots to make better decisions (SAFE, 2011a). Past research has hypothesized that ADM can be taught, and is not merely a by-product of experience (FAA, 1991; Jensen, 1982; "FITS Master," 2006). However, the majority of GA has not yet implemented an effective method of teaching ADM despite a clear need for pilots to improve their ADM skills (SAFE, 2011a).

## **Purpose Statement**

The purpose of this study was to analyze the effectiveness of scenario-based ADM training in improving ADM in collegiate flight students.

## **Hypothesis**

There was a difference in demonstrated ADM between pilots who received scenario-based ADM training and pilots who did not receive scenario-based ADM training, for flight students enrolled in a baccalaureate program at the Daytona Beach campus of Embry-Riddle Aeronautical University (ERAU).

## **Delimitations**

Delimitations for this study included limitations on time and population. The researcher completed the experimental portion of this study entirely within the Fall 2011

semester. The population for this study was limited to ERAU student pilots, solicited from various class sections and student organizations.

### **Limitations and Assumptions**

Budget was a major limitation of the study, limiting the number of participants the researcher could include, and the scope of the treatment the researcher could provide.

Time was another major limitation of the study. Also, the fact that one researcher conducted the entire experiment, including the training and the scoring, made the possibility of bias a limitation of the study.

The self-selected nature of the sample was another limitation of the study, as was the diversity of experience levels in the sample. The original selection criteria for participants limited the participants to those with fewer than 500 hours of total flight time logged and who held at least a private pilot certificate but did not hold any flight instructor certificates. However, the small number of participants who fit the original selection criteria motivated the researcher to include all willing participants. Including all the participants meant accepting a wider range of experience levels in the sample to include student pilots as well as certificated flight instructors (CFIs).

Assumptions of this study included the ability of all parties to understand and communicate effectively in English; English was not the first language for some of the participants but every effort was made to ensure mutual understanding. It was assumed that the researcher was able to accurately assess the participants' decision making throughout the experiment. This study also assumed that the participants answered debrief questions honestly and refrained from discussing the experiment with each other between sessions, as requested by the researcher.

## Definition of Terms

Advanced Aviation Training Device (AATD): a fixed-base flight simulator equipped with full digitally-loaded flight controls, an instrument panel, and a video screen (Frasca International, Inc., 2010).

Aeronautical Decision Making (ADM): “A systematic approach to the mental process used by aircraft pilots to consistently determine the best course of action in response to a given set of circumstances” (FAA, 1991, p. ii).

Active Pilot: A pilot who holds both a pilot certificate and a valid medical certificate issued within the last 25 months (NTSB, 2011a).

Attitude: “A personal motivational predisposition to respond to persons, situations, or events in a given manner that can, nevertheless, be changed or modified through training. A sort of mental shortcut to decision making” (FAA, 1991, p. ii).

Attitude Management: “The ability to recognize hazardous attitudes in oneself and the willingness to modify them as necessary through the application of an appropriate antidote thought” (FAA, 1991, p. ii).

FAA Center For General Aviation Research (CGAR): A consortium of aviation universities conducting research to make significant improvements in safety and efficiency for GA air transportation (CGAR, 2005).

Crew Resource Management (CRM, formerly Cockpit Resource Management): “In multiperson crew configurations, the effective use of all personnel and material assets available to a flight crew. CRM

emphasizes good communication and interpersonal relationship skills” (FAA, 1991, p. ii).

**Decision Process Used:** A metric used to quantify a participant’s aeronautical decision making; describes whether the participant’s actions evidenced a systematic approach in the decision-making process (FAA, 1991).

**FAA Industry Training Standard (FITS):** The FITS program is a collaboration of FAA, industry, and the Center of Excellence for General Aviation. Goals include developing adaptive training and industry standards for the GA community (“FAA-Industry Training Standards [FITS] Program Plan,” 2003).

**Headwork:** Mental work “required to accomplish a conscious, rational thought process when making decisions. Good decision making involves risk identification and assessment, information processing, and problem solving” (FAA, 1991, p. iii).

**Higher Order Thinking Skills (HOTS):** Analysis, synthesis, and evaluation. Levels of cognition which are “essential to judgment, ADM, and critical thinking” (FAA, 2008a, p. 2-5).

**Judgment:** “The mental process of recognizing and analyzing all pertinent information in a particular situation, a rational evaluation of alternative actions in response to it, and a timely decision on which action to take” (FAA, 1991, p. iii).

Margin of Safety: “The difference between pilot capabilities and task requirements” (FAA, 1991, p. 17).

Overall ADM: A numerical score calculated by the researcher to quantify a participant’s aeronautical decision making ability.

Personal Checklist: The checklist of “basic principles that cannot be compromised [including] what not to do” (FAA, 1991, p. 23) The IMSAFE checklist is an example of a personal checklist (FAA, 1991, p. 24).

Personality: “The embodiment of personal traits and characteristics of an individual that are set at a very early age and extremely resistant to change” (FAA, 1991, p. iii).

Poor Judgment Chain: “A series of mistakes that may lead to an accident or incident” (FAA, 1991, p. iii).

Problem-Based Learning (PBL): Lessons structured to confront students with real-world problems and force them to reach realistic solutions by practicing problem-solving skills (FAA, 2008a). Variations of PBL include SBT, collaborative problem-solving, and case study (FAA, 2008a).

Problem Comprehended: A metric used to quantify a participant’s aeronautical decision making analogous to Endsley’s (2000) Level 2 SA, Comprehension; refers to the participant’s ability to integrate “multiple pieces of information and a determination of their relevance to the person’s goals” (Endsley, 2000, p. 4).

**Problem Detected:** A metric used to quantify a participant's aeronautical decision making analogous to Endsley's (2000) Level 1 SA, Perception; refers to the "perception of cues [and] needed information", or the participant's ability to perceive a problem (Endsley, 2000, p. 3).

**Problem Projected:** A metric used to quantify a participant's aeronautical decision making analogous to Endsley's (2000) Level 3 SA, Projection; refers to the ability to "forecast future situation events and dynamics" (Endsley, 2000, p. 4).

**Problem Resolved:** A metric used to quantify a participant's aeronautical decision making; describes whether the participant's reaction to a problem adequately addressed the risk associated with that problem.

**Risk Management:** "The part of the decision making process which relies on situational awareness, problem recognition, and good judgment to reduce risks associated with each flight" (FAA, 1991, p. iii).

**Risk Elements:** "The four fundamental risk elements are the pilot, the aircraft, the environment, and the type of operation that comprise any given aviation situation" (FAA, 1991, p. iii).

**Safe Outcome:** A metric used to quantify a participant's aeronautical decision making; describes whether the participant's reaction to a problem returned the flight to a state in which "the possibility of harm to persons or of property damage is reduced to, and maintained at or below, an acceptable level" (International Civil Aviation Organization [ICAO], 2009, p. 2-2).

Scenario-Based Training (SBT): “A training system that uses a highly structured script of real-world experiences to address flight training objectives in an operational environment” (“FITS Master,” 2006, p. 5).

Situation Awareness (SA, formerly Situational Awareness): “The accurate perception and understanding of all the factors and conditions within the four fundamental risk elements that affect safety before, during, and after the flight” (FAA, 1991, p. iii). Also, “the perception of the elements in the environment within a volume of time and space, comprehension of their meaning and the projection of their status in the near future” (Endsley, 2000, p. 3)

Skills and Procedures: “The procedural, psychomotor, and perceptual skills used to control a specific aircraft or its systems. They are the stick and rudder or airmanship abilities that are gained through conventional training, are perfected, and become almost automatic through experience” (FAA, 1991, p. iii).

Society of Aviation and Flight Educators (SAFE): An organization of aviation educators that works with industry partners and the FAA to provide aviation education resources; goals include fostering professionalism, excellence, and safety (“About SAFE,” 2012).

Stress: “The body’s nonspecific response to demands placed on it, whether those demands are pleasant or unpleasant” (FAA, 1991, p. 17).



Stress Management: “The personal analysis of the kinds of stress experienced while flying, the application of appropriate stress assessment tools, and other coping mechanisms” (FAA, 1991, p. iii).

Timely Manner: A metric used to quantify a participant’s aeronautical decision making; describes whether the participant “execute[d] a suitable course of action within the time frame permitted by the situation” (Jensen, 1982, p. 64).

### **List of Acronyms**

3P	Perceive, Process, Perform
5 Ps	Plan, Plane, Pilot, Passengers, Programming
AATD	Advanced Aviation Training Device
ACT	Aircrew Coordination Training
ADM	Aeronautical Decision Making
AFD	Airport Facility Directory
AOPA	Aircraft Owners and Pilots Association
AQP	Advanced Qualification Program
ASI	Air Safety Institute (formerly the AOPA Safety Foundation)
ATP	Air Transport Pilot
CFI	Certificated Flight Instructor
CFR	Code of Federal Regulations
CGAR	Center For General Aviation Research
CPT	Civilian Pilot Training
CRM	Crew Resource Management

DECIDE	Detect, Estimate, Choose, Identify, Do, Evaluate
DESIDE	Detect, Estimate, Set safety objectives, Identify, Do, Evaluate
DOT	Department of Transportation
ERAU	Embry-Riddle Aeronautical University
FAA	Federal Aviation Administration
FATE	Fly the airplane, Assess the situation, Take action, Evaluate (a Northwest Airlines ADM model)
FITS	FAA-Industry Training Standards
FTD	Flight Training Device
FOR-DEC	Facts, Options, Risks & Benefits, Decision, Execution, Check
GA	General Aviation
GAJSC	General Aviation Joint Steering Committee
GAMA	General Aviation Manufacturers Association
GPS	Global Positioning System
GSC	Guidance, Standardization, and Certification
HAL	High Altitude Lab
HOTS	Higher Order Thinking Skills
ICAO	International Civil Aviation Organization
IMSAFE	Illness, Medication, Stress, Alcohol, Fatigue, Eating
IFR	Instrument Flight Rules
IMC	Instrument Meteorological Conditions
IP	Instructor pilot
LCD	Liquid Crystal Display

LOFT	Line Oriented Flight Training
MBT	Maneuver-Based Training
MFD	Multi-Function Display
NASA	National Aeronautics and Space Administration
NTSB	National Transportation Safety Board
PASS	Problem identification, Acquire information, Survey strategy, Select strategy
PAVE	Pilot in command, Aircraft, enVironment, and External pressures
PBL	Problem-Based Learning
PIC	Pilot In Command
PJ	Poor Judgment
POH	Pilot's Operating Handbook
PTS	Practical Test Standards
RPM	Revolutions Per Minute
SA	Situation Awareness
SAFE	Society of Aviation and Flight Educators
SBT	Scenario-Based Training
SHOR	Stimuli, Hypotheses, Options, Response
SOAR	Situation, Options, Act, Repeat
SRM	Single Pilot Resource Management
TAA	Technologically Advanced Aircraft
UND	University of North Dakota
VFR	Visual Flight Rules

VMC	Visual Meteorological Conditions
VOR	Very High Frequency (VHF) Omnidirectional Radio range

## **Chapter II**

### **Review of the Relevant Literature**

Research into ADM began with developing several key concepts. What eventually became known as ADM was first called judgment. Early research investigated the correlation between judgment and experience, testing the traditional assumption that good judgment developed naturally as a by-product of gaining experience. Other research sought to define expert ADM while still more research identified the component behaviors and mental skills (such as risk management, HOTS, and situation awareness [SA]) associated with expert ADM. Analysis of accident statistics throughout the evolution of ADM training served to indicate how effective different training methodologies were (FAA, 1991).

#### **Judgment, ADM, and HOTS**

Jensen and Benel (as cited in Diehl, 1992) developed a taxonomy of human error that separated decisional task errors from procedural and perceptualmotor errors. They defined decisional errors as errors in mental processes such as planning and evaluation, and emphasized judgment's association with the "complex cognitive processes involved in human decision making" (Diehl, 1992, p. 5). Analysis of aircrew errors from major accidents showed that decisional errors constituted 52%, 56%, and 53% of aircrew errors made in GA, airline, and military accidents respectively (Diehl, 1992).

Researchers later began referring to decisional tasks as judgment, or decisional judgment (Jensen, 1982; Diehl, 1992). Jensen (1982) presented a working definition of judgment that applied to aviation:

(1) The ability to search for and establish the relevance of all available information regarding a situation, to specify alternative courses of action, and to determine expected outcomes from each alternative.

(2) The motivation to choose and authoritatively execute a suitable course of action within the time frame permitted by the situation, where: (a) “Suitable” is an alternative consistent with societal norms; (b) “Action” includes no action, some action, or action to seek more information. (p. 64)

Jensen’s definition of judgment described it as a combination of many complementary mental functions and incorporated both cognitive and motivational components. Other common terms for this combination of mental functions included “headwork” and “staying ahead of the aircraft” (Jensen, 1982). The FAA has since defined judgment as “the mental process of recognizing and analyzing all pertinent information in a particular situation, a rational evaluation of alternative actions in response to it, and a timely decision on which action to take” (FAA, 1991, p. iii).

Over time, ADM became the more common term used to describe these mental tasks. The FAA defined ADM as “a systematic approach to the mental process used by aircraft pilots to consistently determine the best course of action in response to a given set of circumstances” (FAA, 1991, p. ii). The FAA’s definition of ADM shared many characteristics with Jensen’s earlier definition of judgment, although the FAA’s definition simplified Jensen’s (1982) itemized list of mental tasks as “the mental process” (FAA, 1991, p. iii). An understanding of this “mental process” was developed over many years of research.

Bloom's research on the cognitive domain of learning provided valuable insight into the mental processes associated with ADM. Specifically, Bloom's taxonomy of the cognitive domain gave aviation researchers an accurate, theoretical description of those mental processes (FAA, 2008a). Researchers refined the definition of ADM using the taxonomy's more complex levels of thinking, known as the HOTS. Bloom's taxonomy described six levels of thinking behaviors that progressed along a continuum from simple to complex: knowledge, comprehension, application, analysis, synthesis, and evaluation (FAA, 2008a). Analysis involved recognizing, examining, and understanding information from the environment. Synthesis involved combining information into a new and integrated whole. Evaluation involved judging the benefits and disadvantages of an idea or phenomenon (FAA, 2008a). HOTS were so essential to judgment and decision-making that the FAA used HOTS to partially define ADM for aviation instructors (FAA, 2008a).

### **SA (formerly Situational Awareness)**

Research into the decision-making process included investigations into a related mental process known as SA (formerly situational awareness). Endsley (2000) defined SA as "the perception of the elements in the environment within a volume of time and space, comprehension of their meaning and the projection of their status in the near future" (p. 3), or more simply as "knowing what is going on around you" (p. 2). Her *Theoretical Underpinnings of Situation Awareness: A Critical Review* provided an overview of the SA construct and associated terms (Endsley, 2000).

Endsley (2000) explained that making a decision, like any other task, was enabled by accurate SA. The relationship between SA and ADM was more probabilistic than a

direct correlation. High SA increased the probability of successful ADM, but did not guarantee it (Endsley, 2000).

Endsley (2000) divided SA into three levels, with each level enabling the next. Level 1 SA, perception, involved distinguishing important information from the environment via a sensory organ (Endsley, 2000). Level 2 SA, comprehension, involved combining and interpreting perceived information to derive meaning about the current situation (Endsley, 2000). Level 3 SA, projection, described the ability to predict the future state of a situation based on an understanding of the current situation (Endsley, 2000).

Endsley and Garland (2000) reported that SA was a “considerable challenge in [GA] as GA pilots are frequently less experienced and less current than operators for major airlines” (p. 357). They observed that a common SA error in low experience GA pilots was a tendency to overestimate their skill level and underestimate the severity or risk of a situation. There is a need to improve GA pilots’ SA, as Endsley and Garland (2000) discussed in their paper, *Pilot Situation Awareness Training in General Aviation*.

### **GA Pilots**

GA pilots need improved ADM training in order to improve GA safety. GA includes all aviation except military and scheduled commercial operations (General Aviation Manufacturers Association [GAMA], 2011). GA pilots fly aircraft ranging from two-seat trainers to long-range jets. GA pilots have varying levels of certifications and hours logged, and generally do not receive as much recurrent training as military or commercial pilots (Endsley & Garland, 2000). GA also serves as the main training environment for future commercial airline pilots (GAMA, 2011).



GA suffered the highest accident rates in civil aviation in the last decade (NTSB, 2012). The overall accident rate remained at around six accidents per 100,000 flight hours, and the fatal accident rate remained at around one accident per 100,000 flight hours (NTSB, 2011a). By comparison, accident rates for Part 121 operations continuously improved over the last decade and dropped to 0.152 and 0.006 accidents per 100,000 flight hours for total and fatal accidents, respectively, in 2009 (NTSB, 2011b). When the NTSB added GA Safety to its *Most Wanted List*, it noted that the causes of current GA accidents continued to repeat the causes of historical GA accidents (NTSB, 2012).

Meanwhile, the GA pilot population has been shrinking. There were 532,177 active pilots in 2000, compared to 494,177 active pilots in 2011 (NTSB, 2011a). The number of pilot certificates issued annually declined for all certificate categories. For example, private pilot certificates issued declined from 27,223 in 2000 to 13,457 in 2010 (General Aviation Manufacturers Association [GAMA], 2011). Commercial certificates issued declined from 11,813 in 2000 to 5,774 in 2010 (GAMA, 2011).

The NTSB noted an overall decline in the number of hours flown in GA since 2000 (NTSB, 2011a). There was a sharp decline in hours flown from 2002 to 2003, likely because of the restrictions imposed on GA after the terrorist attacks in 2001. Whatever the reason, hours flown never recovered to pre-2001 levels and dropped sharply again from 2007 to 2009, most likely as a result of economic factors (NTSB, 2011a). Data collected by GAMA shows that the downward trends in GA activity since 2000 were a continuation of negative trends begun in preceding decades (2011).

**ERAU student pilots.** Enrollment at Embry-Riddle Aeronautical University's Daytona Beach campus was 4,496 in fall of 2010 ("Enrollment," 2012). The student body included 1,101 students enrolled in the B.S. of Aeronautical Science degree program and 275 enrolled in the B.S. of Aeronautics degree program; these two degree programs included all of ERAU's flight students ("Enrollment," 2012). International students from 99 foreign countries made up 14% of the Daytona Beach campus population ("Student Demographics," 2012). The average age of students at the Daytona Beach campus was 21 years old, although some were as young as 16 and many students were in their 20s and 30s (the population included many veterans, for example) ("Student Demographics," 2012). The population was mostly male, as is typical in the aviation industry; only 17% of the residential campus students were female ("Student Demographics," 2012).

### **Expert ADM**

Tradition associated good judgment so strongly with experience that aviation researchers based their definitions of good ADM on the ADM exhibited by an experienced – or expert – pilot.

**The novice pilot.** To define expert ADM, researchers first had to distinguish between novice and expert pilots. The term *novice pilot* has been used differently by various researchers to describe a relatively inexperienced pilot. Beringer and Schvaneveldt (2002) categorized pilots as novice or experienced by using overall flight time as a measure of experience, which is a traditional but flawed measurement. Deitch (2001) used *novice pilot* and *student pilot*, a person training to become a private pilot, synonymously. Kobus, Procter, Bank, and Holste (2000) defined *novice* and *expert* by

measuring their population sample against itself; the median experience level was used as the break point between *novice* and *expert*.

Other researchers distinguished novices from experts based on mental capabilities. Wiggins and O'Hare (2003) distinguished novice pilots from expert pilots according to "individual differences [...] in their capacity to recognize and respond appropriately to deteriorating [...] conditions" and further qualified expert pilots by saying "experts outperform novices in the capacity to acquire information" (p. 337-338). Endsley and Garland (2000) described differences in SA between groups of pilots with different experience levels. They compared GA pilots (approximately 720 hours experience) to airline pilots (approximately 6,000 hours experience) and to commercial airline check airmen (approximately 12,000 hours experience). They reported that more experienced pilots demonstrated increasing levels of preflight preparation and more focus on understanding and projection – more effective SA that enabled more successful ADM. Endsley and Garland concluded that SA training that addresses SA problems typical of low-experience GA pilots should be effective at improving SA and therefore ADM.

**The accident-prone pilot.** Understanding expert ADM also required research on faulty ADM, such as the ADM that causes a pilot to have an accident. Adams, Hamilton, Koonce, and Hwoschinsky (2002) completed a study that analyzed surveys from 4,000 pilots to develop an index that could classify a pilot as high-risk or accident-prone. Their index was successful at predicting whether a pilot within their population sample had an accident or not (their population sample included pilots who had an accident and control pilots whose records were accident-free). In developing the index, they were able to characterize the ADM styles of both accident-prone and accident-free pilots. They

concluded that high-risk or accident-prone pilots were more likely to “expose themselves to unsafe flying experiences, feel time pressure when making decisions, have a false sense of their ability to handle the situation, and not review alternative options or solutions” (Adams et al., 2002, p. 948).

An NTSB review of GA accidents led to a profile of GA pilots most likely to have accidents (Endsley & Garland, 2000). The greatest number of accidents involved “pilots between 35 to 39 years of age with between 100 to 499 hours total time who were engaged in personal flying” (Wells, 1992, as cited in Endsley & Garland, 2000, p. 2). Within the period from 100 to 499 hours total time, pilots with about 100 hours total time or who were 50 to 100 hours beyond the private-pilot certification or instrument-rating were particularly accident-prone. Pilots who were recently certificated tended to overestimate their capabilities and put themselves in riskier situations (Trollip & Jensen, 1991, as cited in Endsley & Garland, 2000). Endsley and Garland concluded that GA pilots who fit this profile were particularly receptive to, and especially needful of, specialized SA and ADM training.

Although the typical accident-pilot profile described a low-time private pilot who may or may not hold an instrument rating, pilots with additional flight time and certificates made up a significant percentage of accident pilots. Of the non-commercial fixed-wing GA accidents in 2009, 24% involved commercial pilots and 13% involved ATPs (AOPA, 2010). A CFI was on board in 21% of those accidents (AOPA, 2010). Lethality of accidents was relatively constant for all levels of pilot certification involved in non-commercial fixed-wing accidents, with the exception of student pilots whose lethality rate was about one-quarter that of the other certification levels (AOPA, 2010).

Of the non-commercial helicopter GA accidents in 2009, 63% involved commercial pilots, 13% involved ATPs, and a CFI was on board in 51% of those accidents (AOPA, 2010).

### **ADM Training**

Using expert ADM as the goal for new ADM training programs, researchers examined traditional ADM training methods. Researchers developed new ADM training programs that focused on behavior management and on a combination of problem-solving skills and practice.

**Traditional ADM training: informal by-product of experience.** Aviation first realized the extent of pilots' weaknesses in decision-making, communication, and coordination when cockpit voice recorders and flight data recorders were first used in accident investigations in the 1970s (Diehl, 1992). This discovery prompted the FAA and industry to develop formal decision-making training programs. At the time, civilian flight training remained unchanged from the CPT program implemented in advance of World War Two (Wright, 2009). The CPT training program was maneuvers-based with eligibility for certification defined by performance of those maneuvers within minimum standards and accumulation of minimum amounts of training time. Completion of training meant passing the FAA's knowledge and practical tests (Diehl, 1992). Judgment was expected to develop as a natural by-product of experience after the check ride (Jensen, 1982). Many studies revealed a correlation between experience and higher quality decision-making (Schrivver, Morrow, Wickens, & Taulleur, 2008).

Adams (1992b) explained how ADM was thought to develop in traditional flight training. Traditional training through experience enabled pilots to develop problem-

solving ability first by applying rote procedures to handle a situation that had been covered in training (Adams, 1992b). Novice pilots then developed a store of procedural knowledge from encountering real-world problems and operational constraints (Adams, 1992b). After 1,000 to 10,000 hours, an expert pilot could apply responses quickly, based on similar past experiences and could begin to integrate knowledge learned from past experiences into solutions to solve novel situations (Adams, 1992b).

Despite a correlation between improved ADM and increasing flight experience, experience alone had not been proven as the most effective method for acquiring judgment. Scholarly research and accident statistics showed that a significant percentage of accidents involved pilots with higher certifications and more experience (AOPA, 2010). Training ADM informally as a by-product of experience was no longer adequate, as evidenced by accident investigations and statistics (Diehl, 1992).

Industry researchers began developing formal ADM training programs. Gaining judgment through experience required time, money, and exposure to the very situations in which pilots need good decision-making skills to maintain safety of flight (Molesworth, Wiggins, & O'Hare, 2006). As Jensen (1982) stated, without decision-making training, "it is but a slight overstatement to say that good pilot judgment is learned by the lucky and the cautious over many years of varied flying experience" (p. 61).

**ADM: behavior management.** Formal ADM training in all segments of aviation began from the theoretical foundations of behavior management and management theory. Early research on ADM emphasized changing pilot attitudes (Kochan, Jensen, Chubb, & Hunter, 1997). Researchers thought of faulty ADM as a result of misplaced motivation or a psychological factor in the pilot. GA ADM training materials therefore focused on

making pilots aware of hazardous attitudes and management of stress and risks (Kochan et al., 1997).

In the late 1980s and early 1990s, the FAA published six manuals and an advisory circular to provide official guidance on GA ADM training. These documents were based on more than twelve years of research, development, and testing, and they represented the first formal effort to provide guidance on formal ADM training (FAA, 1991). These materials became the standard on which ADM training was based for a significant period of time.

The focus on behavior management was consistent and evident throughout. For example, the ADM manual for student and private pilots called *Aeronautical Decision Making for Student and Private Pilots* stated that its purpose was to explain the risks associated with flying, the behavioral causes of typical accidents, and the impact of stress on decision making (Diehl, Hwochinsky, Lawton, & Livack, 1987).

The remaining FAA manuals were aimed at GA pilots at different levels of training. Besides student and private pilots, the other manuals were designed for instrument pilots, commercial pilots, instructor pilots, helicopter pilots, and pilots working in multi-pilot crew environments (Adams & Thompson, 1987; Bush, Lawton, & Livack, 1987; Jensen, 1989; Jensen & Adrion, 1988; Jensen, Adrion, & Lawton, 1987). Each rating-specific manual explained the risks associated with that specific type of flying activities. They then described the “underlying behavioral causes” of judgment error and the effects of stress on decision making. They emphasized managing the pilot’s behavior and stress as a way to avoid unnecessary risk (Adams & Thompson, 1987; Bush et al., 1987; Diehl et al., 1987; Jensen, 1989; Jensen & Adrion, 1988; Jensen et al., 1987).

Teaching exercises included helping the student assess their own hazardous attitudes and identifying the hazardous attitudes exhibited by pilots who had accidents (Adams & Thompson, 1987; Bush et al., 1987; Diehl et al., 1987; Jensen, 1989; Jensen & Adrion, 1988; Jensen et al., 1987). Finally, the manuals directed instructors to teach better judgment by exposing students to flight situations drawn from actual accidents and incidents, asking the students for input, and then giving feedback on the students' responses (Jensen, 1989).

*Aeronautical Decision Making* began by stating that “good judgment can be taught” (FAA, 1991, p. 1). The advisory circular's stated purpose was to provide a “systematic approach to risk assessment and stress management in aviation, illustrate how personal attitudes can influence decision making and how those attitudes can be modified to enhance safety” (FAA, 1991, p. i). The document described hazardous attitudes and stress before outlining exercises to help students identify hazardous attitudes in others and assess themselves.

The advisory circular provided a list of ADM definitions that have become industry standard (FAA, 1991). The advisory circular also described the DECIDE model, a six step, continuously looping process intended to give pilots a logical way to approach decision making. The six steps of DECIDE were: detect, estimate, choose, identify, do, and evaluate (FAA, 1991). The IMSAFE Checklist was offered as a method for assessing the risks associated with a pilot's personal state. IMSAFE stood for illness, medication, stress, alcohol, fatigue, and eating (FAA, 1991).

The FAA also provided some specific aspects of ADM instruction. As with previous guidance, this document advocated discussing scenarios with the student to



ensure the student understood the hazardous attitudes (FAA, 1991). There was a new emphasis on the flight instructor's role in ADM training (a combination of role model, evaluator, and coach). The FAA said students needed to be exposed to ADM instruction earlier, and instructors needed to teach ADM in the air as well as on the ground. The flight instructor needed to create in-flight scenarios to "stimulate the student's decision making process" and respond to student behavior in a way that encourages safe decision making (FAA, 1991, p. 28).

The effectiveness of the FAA's manuals was validated in multiple independent studies where student pilots received training in accordance with the manuals along with a standard flying curriculum (FAA, 1991). Pilots in empirical studies made significantly fewer in-flight errors after receiving ADM training; the reduction in judgment errors ranged from 10% to 50% (FAA, 1991). Pilots in an operational study at Petroleum Helicopter Inc. demonstrated a 54% reduction in overall accident rate after receiving recurrent training in accordance with the FAA's ADM manuals (Diehl, 1992; FAA, 1991).

Although the effectiveness of the ADM training described in the FAA's guidance was validated by several studies, formal ADM training was not effectively integrated into GA training. The FAA had provided guidance on ADM training but did not alter the testing standards or certification methods to clearly define satisfactory ADM. Satisfactory performance continued to be defined in terms of minimum knowledge, proficiency, and aeronautical experience (Wright, 2002). With no clear standard for certification and no motivation, GA training continued to "teach to the test" instead of using formal ADM training.

**ADM: problem-solving skills and practice.** Accident rates in GA plateaued during the 1990s (Wright, 2002). The ultimate goal of ADM training was to improve GA safety by reducing the occurrence of decision error accidents. Researchers interpreted the lack of improvement in accident rates as an indication that current ADM training was having little measurable impact on GA safety. This conclusion led to analyses of the current ADM training materials and investigations into how to improve and better implement ADM training methodologies. Research about ADM began to focus on ADM as a problem-solving skill set, and began to emphasize practicing problem-solving in ADM training instead of attitude management (Adams, 1992a; Adams, 1992b; Irving, 1992; Ericsson, 1992).

Many researchers presented their findings on existing ADM training at the FAA's ADM Workshop in 1992 (Adams & Adams, 1992). Adams (1992b) described the judgment that expert pilots exhibited as composed of a "variety of different processing skills and unique problem solving capabilities" (p. 110). He criticized the applicability of the ADM training manuals developed by the FAA described above, saying that they taught an algorithmic, linear process of methodical decision-making which did not resemble the way experts actually made decisions when in emergency or stressful conditions (Adams, 1992b).

Adams (1992b) suggested several alternative training methods. Activity-based learning would engage flight students in real-world problems while in the training environment, allowing students to gain experience in a controlled manner. Exercises and discussions on SA based on vignettes would help students practice maintaining their SA

(Adams, 1992b). Interactive computer or video training devices would be useful in creating realistic activities (Adams, 1992b).

In another presentation, *How Expert Pilots Think*, Adams (1992a) revisited the role of experience in developing problem-solving skills in contemporary ADM training. He observed that practice was the most important variable in determining the level of expertise an individual achieved in non-aviation domains (Adams, 1992a). Flying experience was important in helping develop a pilot's base of knowledge and procedural responses in that the practice enabled faster pattern recognition, problem perception, and more efficient problem solving (Adams, 1992a). Real-life flying rarely presented opportunities to practice problem-solving, though, which suggested that pilots need ADM training that provides more opportunity to practice those skills in order to gain expertise (Adams, 1992a).

Irving (1992) also emphasized ADM as a complex problem-solving process. He criticized the traditional ADM training method, which consisted primarily of on-the-job training. Such training was restricted to the type of normal, day-to-day occurrences that did not require advanced ADM skills. Random instead of structured, traditional ADM training was not formalized in such a way as to ensure all the important skills were covered (Irving, 1992).

Irving (1992) also criticized first-generation ADM training for relying too heavily on formulaic procedures. Pilots were being trained to simply apply the procedure instead of exploring alternative solutions (Irving, 1992). Performance was measured through observation of easily measured motor skills and on how accurately the procedure was applied, rather than valid measurements of ADM (Irving, 1992). Programs provided little

or no guidance in acquiring data acquisition skills (Irving, 1992). Finally, training was not being provided by expert decision-makers or based on input from subject matter experts (Irving, 1992).

Irving (1992) suggested that ADM training programs incorporate structured scenarios instead. Effective scenarios had to be realistic, ideally drawn from observations of more experienced pilots (Irving, 1992). Scenarios would establish training objectives beforehand and include a full debriefing afterwards (Irving, 1992). Using scenarios would create opportunity for practicing the process of evaluating and solving dangerous situations and review alternate solutions while safely in a hangar (Irving, 1992).

Ericsson's (1992) presentation, *Methodology for Studying and Training Expertise*, echoed the points made by Adams (1992a; 1992b) and Irving (1992). Ericsson (1992) described expert decision making in aviation as a set of critical skills including evaluation, correlation, and application of relevant memories. Deliberate practice was necessary to acquire and maintain expert performance in aviation, just as in other domains where expert performance was observed such as competitive sports and medicine (Ericsson, 1992).

The researchers who gathered for the *Workshop on Aeronautical Decision Making* formulated an action plan to improve ADM training effectiveness (Adams & Adams, 1992). The plan identified several key participants within the federal government, industry, and academia (Adams & Adams, 1992). The plan identified four major tasks that needed to be accomplished in order to improve overall effectiveness of ADM training. Those tasks were to define the structure of decision making tasks, develop

training requirements, specify training strategies, and evaluate training effectiveness (Adams & Adams, 1992).

**Problem-based learning.** Several ADM training strategies were created and tested after the *Workshop on Aeronautical Decision Making* (Adams & Adams, 1992). The conceptualization of ADM as a set of problem-solving skills (rather than the result of a behavior or motivation problem) led many aviation researchers to investigate adapting problem-solving strategies that had been developed in other fields to aviation. One such strategy was called problem-based learning (PBL).

Medical researchers at McMaster University School of Medicine pioneered the PBL approach to teaching and curriculum design in 1966 (FAA, 2008). PBL was defined as lessons structured to confront students with real-world problems and force them to reach realistic solutions by practicing problem-solving skills (FAA, 2008). Variations of PBL included SBT, collaborative problem-solving, and case study (FAA, 2008). SBT used a highly structured script based on real-world experiences to “address aviation training objectives in an operational environment” (FAA, 2008, p. 4-16). Collaborative problem-solving engaged multiple students in collaborative problem-solving discussions guided by an instructor (FAA, 2008). In case study training, the instructor presented an account of a real world situation that illustrated a point and then prompted the students to analyze the case, develop possible solutions, and come to conclusions (FAA, 2008).

***Commercial and military training implementation.*** Forms of problem-solving learning were adopted by the commercial and military aviation sectors after attending a 1979 workshop hosted by the National Aeronautics and Space Administration (NASA). Participants learned that the majority of human errors that led to accidents were failures

of decision making, leadership, and communications (Helmreich, Merritt, & Wilhelm, 1999). Commercial airlines and military leadership left the workshop determined to create training programs to prevent these errors. Such programs were strikingly popular and evolved rapidly in an innovative and collaborative environment (Helmreich & Foushee, 2010).

Commercial airlines created training programs to “enhance interpersonal aspects of flight operations” (Helmreich et al., 1999, p. 1). These programs were known as crew resource management (CRM) training programs (Diehl, 1992). United Airlines developed the first comprehensive U.S. CRM program (Helmreich et al., 1999). KLM developed a leadership training program, while Northwest pioneered Line Orientated Flight Training (LOFT), a form of SBT which modeled each training session after a real-life, or “line”, flight (Diehl, 1992). Most CRM programs evolved to include training manuals, interactive classroom discussions, and LOFT sessions (Diehl, 1992).

CRM programs proved to be very effective at reducing pilot error in air carrier operations. A case-series analysis of crashes and other mishaps of domestic air carrier flights (operating under Title 14 Code of Federal Regulations [CFR] Part 121), both scheduled and nonscheduled, that occurred during 1983 – 2002 revealed several encouraging trends (Baker, Qiang, Rebok, & Li, 2008). The proportion of mishaps involving pilot error decreased from 42% in 1983-1987 to 25% in 1998-2002 (Baker et al., 2008). Mishap rates related to poor decision making decreased from 6.2 to 1.8 per 10 million flights, and mishap rates involving poor crew interaction declined from 2.8 to 0.9 per 10 million flights (Baker et al., 2008). Baker et al. (2008) credited these

improvements to air carriers' emphasis on CRM as well as improving technology such as cockpit displays.

The U.S. Air Force Military Airlift Command (now Air Mobility Command) and the U.S. Naval Safety Center pioneered Aircrew Coordination Training (ACT), the military equivalent of CRM programs (Diehl, 1992). ACT programs were designed to improve decision making as well as communications within military cockpits and between crews and outside contacts (O'Conner, Hahn, & Nullmeyer, 2010). ACT programs remained largely unchanged in the 1980s, but military-funded research in the early 1990s led to advances in CRM training effectiveness (O'Conner et al., 2010). By 2010, the U.S. Navy, Marine Corps, Army, Air Force, and Coast Guard all utilized CRM training programs (O'Conner et al., 2010). Many non-U.S. military services had implemented a CRM program by 2010 too (O'Conner et al., 2010). Recent research evaluated the effectiveness of SBT-based CRM training in the People's Republic of China Air Force (Li & Harris, 2005, 2008).

Li and Harris published a study in 2005 that evaluated the suitability of various ADM mnemonics for resolving different types of decision-making scenarios. The researchers asked instructor pilots (IPs) in the Chinese Air Force Academy to rate the suitability of the following ADM methods: SHOR (Stimuli, Hypotheses, Options, Response), PASS (Problem identification, Acquire information, Survey strategy, Select strategy), FOR-DEC (Facts, Options, Risks & Benefits, Decision, Execution, Check), SOAR (Situation, Options, Act, Repeat); and DESIDE (Detect, Estimate, Set safety objectives, Identify, Do, Evaluate). The IPs favored two of the mnemonics depending on how much time was available to make a decision; the IPs judged SHOR to be the best

method for time-limited decisions while DESIDE was deemed more suitable for decisions that were less time-limited and required more comprehensive thinking (Li and Harris, 2005).

Li and Harris (2008) later created and tested an ADM training program that they administered to a group of Chinese Tactical Training Wing pilots. Half of the participants received Li and Harris' ADM training course while the other half did not. All of the participants then completed simulated flights in a full-flight simulator where the participants' decision-making skills were evaluated with respect to situation assessment, risk management, and response time. Those pilots who received the ADM training exhibited significant improvements in the quality of their situation assessment and risk management, although response time was negatively impacted. Li and Harris (2008) concluded that their ADM training program was effective in improving decision making and that ADM was trainable.

***GA training research.*** Several forms of PBL were tested in GA settings as well. O'Hare, Mullen, and Arnold completed a study in 2009 testing the effectiveness of case-based reflection. O'Hare et al. (2009) gathered a sample of non-pilots, and provided different ADM training to groups of test subjects. All of the subjects read case studies where a pilot encountered adverse weather. Half of the subjects read cases where the pilot successfully dealt with the conditions and landed safely while the other half read cases where the pilot crashed. After reading the case studies, half of the subjects participated in a reflective thinking exercise while the other half merely recalled as much detail as possible about the cases. They then flew a simulated flight on a computer-based flight



simulator and had to decide when (or if) to discontinue a flight as the researchers gradually made the weather conditions deteriorate (O'Hare, et al., 2009).

Those participants who reflected on the cases stopped the flight sooner, when the weather had not deteriorated as far, than those who merely recalled the cases (O'Hare, et al., 2009). Several participants who merely recalled the cases failed to discontinue the flight and crashed into terrain. The outcome of the cases did not have a significant impact on the participants' decision making. These results led the researchers to conclude that reflecting on cases improved ADM with respect to recognition of critical weather situations and adherence to relevant regulations. Whether the cases studied resulted in success or a crash did not seem to have any significance. Although the researchers intentionally selected participants who were not pilots, the apparent improvement in decision-making as a result of reflection demonstrated the utility of case-based training for ADM (O'Hare et al., 2009).

Lee, Fanjoy, and Dillman (2005) examined the effects of regular exposure to safety information on the ADM capacity of students in a collegiate flight program. The study focused on ADM involved in mechanical malfunction scenarios. The researchers took three measurements of the participants' ADM: recognition time, response time, and appropriateness of response (Lee et al., 2005). The population consisted of undergraduate students who had received their private pilot certificate and were training for a commercial pilot certificate. The experimental group of students received online access to a safety information system that compiled aircraft discrepancies; the experimental group also received online prompts to review the information in the safety information system before each of their routine training flights over a five-week period. The results showed a

measurable improvement in recognition time, response time, and appropriateness of response in experimental participants compared to the control participants (Lee et al., 2005). The study findings therefore supported the hypothesis that regular exposure to safety information improves ADM (Lee et al., 2005).

**ADM mnemonics and acronyms.** Operators used many mnemonics and acronyms to help pilots remember CRM, ADM, and Single Pilot Resource Management (SRM) concepts. Some examples included FATE, which summarized the basic steps in the ADM process: Fly the aircraft; Assess the situation; Take appropriate action; Evaluate the results (Sumwalt & Watson, 1995). The FAA developed other mnemonics for ADM including the Three-P (3P) model. According to the 3P model, the pilot applied ADM by *perceiving* the current flight circumstances, *processing* the significance of those circumstances, and *performing* the best course of action [emphasis added] (FAA, 2008a).

Other ADM mnemonics included SHOR, PASS, FOR-DEC, SOAR, and DESIDE (Li & Harris, 2005). The FAA promoted the IMSAFE and PAVE checklists to help pilots manage risk (FAA, 2008a). IMSAFE evaluated personal risk factors, as described previously; PAVE divided flight risks into four categories of “Pilot in command, Aircraft, enVironment, and External pressures” (FAA, 2008a, p. 9-6). The Five Ps (5 Ps) was another commonly used memory aid for evaluating the risk of a flight. The 5 Ps consisted of “the Plan, the Plane, the Pilot, the Passengers, and the Programming” (FAA, 2008a, p. 9-13).

### **Implementing GA ADM Training**

Formal ADM training still had not been widely implemented in GA as recently as 2009 although PBL-based ADM training programs were implemented decades earlier by

commercial and military operators (Wright, 2009). Implementation of CRM, LOFT, and Advanced Qualification Program (AQP) at commercial carriers resulted in significant improvements in decision-related accident rates and in overall pilot error rates (Baker et al., 2008; Wright, 2009). In contrast, accident statistics show that pilot-error accident rates for GA improved very little during the 1990s and remained unchanged during the past decade (AOPA, 2009; AOPA, 2010).

The lack of measurable improvement in decision error rates and in overall accident rates in GA prompted escalating responses from the FAA, industry, and academia from the late 1990s on. The FAA formed CGAR and founded the FITS program. These programs involved increasingly collaborative efforts between the FAA and industry representatives such as manufacturers and operators, and educational institutions to develop consensus-based standards. The FAA also published revised guidance on ADM training – emphasizing the use of SBT – in new versions of several manuals and practical test standards (PTS) between 2008 and 2011 (FAA, 2008a; FAA, 2008b; FAA, 2010a; FAA, 2010b; FAA, 2011a; FAA, 2011b).

**FAA CGAR.** The FAA founded CGAR in 2001 (CGAR, 2011). Aviation universities including ERAU, University of Alaska, University of North Dakota (UND), and Wichita State University coordinated through CGAR to support industry and FAA research goals. CGAR’s mission is to “make significant contributions toward improvements in safety and efficiency for GA air transportation [...] with multidisciplinary teams to enhance aviation related research, education, technology transfer and the utilization of research in mission critical areas” (CGAR, 2011, p. 1).

**FITS.** The FAA launched the FITS program under the *Safer Skies* program in 2002 (“FAA-Industry Training Standards [FITS] Program Plan,” 2003). There was growing support to “train the way you will fly (in the real world) and fly the way you were trained” in GA (Wright, 2002, p. 10). The overall goals of the FITS program were to identify changing training needs and develop standards based on industry consensus that responded to the pace of development in GA (“FAA-Industry Training Standards [FITS] Program Plan,” 2003). FITS aimed to create scenario-based, learner-focused training materials that would produce pilots with more practical knowledge and skills than traditional training provided (“FAA-Industry Training Standards [FITS] Program Plan,” 2003). Supporting goals included developing a new GSC infrastructure to support reforms in GA training (“FAA-Industry Training Standards [FITS] Program Plan,” 2003).

Initial members of FITS included the FAA, CGAR, and industry leaders such as Eclipse Aviation, Adam Aircraft, Cessna Aircraft, and Elite Air Center (“FAA-Industry Training Standards [FITS] Program Plan,” 2003). An oversight committee composed of FAA and industry members oversaw the FITS program plan, goals, methodology, schedule, and tasking (“FAA-Industry Training Standards [FITS] Program Plan,” 2003). The FAA also created a FITS Work Group which evaluated the products FITS delivered, and developed recommended training programs and guidance (“FAA-Industry Training Standards [FITS] Program Plan,” 2003).

The *FAA-Industry Training Standards [FITS] Program Plan* (2003) explained reasons to reform the existing GSC in GA. Existing GSC was comprised of advisory circulars, handbooks, PTS, and other materials such as the ADM manuals described above (Adams & Thompson, 1987; Bush et al., 1987; Diehl et al., 1987; Jensen, 1989;

Jensen & Adrion, 1988; Jensen et al., 1987). FITS researchers doubted whether many GA training operators used the existing GSC at all (“FAA-industry,” 2003). Many documents were so out-of-date as to be obsolete (“FAA-industry,” 2003). Revising these documents was a lengthy process and the FAA had no method for managing the GSC’s currency (“FAA-industry,” 2003). The accelerating pace at which technology was modernizing GA aircraft, navigation, and airspace only aggravated GSC’s inflexibility. Other GSC material was incomplete. The GSC material that was current was oriented towards teaching to the knowledge and practical tests rather than developing ADM, SA and other HOTS through an SBT and performance-based testing approach (“FAA-industry,” 2003). As it was, the GSC prevented a reform of the GA training paradigm because it maintained the current maneuver-based training (MBT) paradigm (Wright, 2002).

The overall goals of the FITS program were to identify changing training needs and develop industry standards that responded to the pace of development in GA and were based on industry consensus (“FAA-industry,” 2003). Supporting goals included developing a new GSC infrastructure to support reforms in GA training (“FAA-industry,” 2003). FITS was not intended as a regulatory mechanism, rather, the goal has been to create safer pilots in less time by creating voluntary alternatives to regulatory-mandated training (Glista, 2003). Possible applications included an FAA-approved proficiency program and aircraft or equipment specific training that would lower insurance premiums (Glista, 2003).

Researchers at FITS universities conducted studies to investigate the effectiveness of FITS training. A study conducted at ERAU in the fall of 2004 compared collegiate students training towards an instrument rating in traditional MBT to students who were

instructed using SBT (French, Blickensderfer, Ayers, & Connolly, 2005). The population sample was 27 ERAU students training for the instrument rating. The researchers randomly assigned the participants to the control (MBT) or experimental (SBT) groups. All participants received eight hours of training before the final evaluation. An experimentally blind rater evaluated the participants' instrument flight skills using a computer-based flight simulator program on pre- and post-training "data collection flight[s]" (French et al., 2005). Both MBT and SBT trained students showed significant improvements between the pre- and post-test measures. Furthermore, the SBT group performed statistically better on many measures of piloting ability than the MBT group in the post-test measures. The results suggested that SBT may improve instrument flight rules (IFR) piloting and navigation skills over traditional MBT in a TAA aircraft (French et al., 2005).

UND conducted a study that compared the effectiveness of SBT in teaching HOTS and decision making to traditional aviation instruction and self-study (Robertson, Petros, Schumacher, McHorse, & Ulrich, 2006). The study used 45 undergraduate UND students, divided into three groups (Robertson et al., 2006). All participants were upper-level undergraduate students who were qualified to fly at least one other single-engine piston aircraft (Robertson et al., 2006). The SBT-trained group was trained using transition training that UND researchers had previously designed for Cirrus Aircraft to transition pilots into the Cirrus SR22; PBL was incorporated into both ground and flight instruction (Robertson et al., 2006). The self-study group used the *Cirrus SR 22 Pilot's Operating Handbook* (POH) and CD-ROM-based training materials; their instruction consisted of being presented with scenarios and asked to research the POH to find a

solution for the scenarios (Robertson et al., 2006). The alternate treatment group was used as a control and received non-PBL instruction and MBT similar to traditional transition training (Robertson et al., 2006).

Pre- and post-training sessions conducted in a simulator provided the data to evaluate the participants' aircraft control, pilot performance, SA, and ADM (Robertson et al., 2006). Eight research assistants either possessed a CFI certificate or received training to conduct the evaluations. The research assistants also provided the ground and flight training, then evaluated the students that the other assistants had trained (Robertson et al., 2006). The results indicated significant improvements in measurements of judgment and ADM, SA, and automation management ability (Robertson et al., 2006). None of the differences were statistically significant, though, so the study had difficulty supporting a shift in GA training from MBT to SBT (Robertson et al., 2006). The study did not identify any weaknesses of FITS training as implemented in the study (Robertson et al., 2006).

The FITS program's emphasis on higher-order skills such as enhanced decision making and single-pilot resource management represented a significant departure from the established MBT philosophy. As Wright (2011) discussed, the intention behind FITS was to reform the old training paradigm to better train GA pilots to deal with current safety issues. FITS research, such as the studies discussed above, validated the effectiveness of SBT in GA training and the effectiveness of collaboration between the FAA and industry. Other FITS studies verified that FITS SBT training was high quality and participants were very satisfied (Robertson & Summers, 2007). Despite the FITS program's validation of SBT, progress on reforming the GA training paradigm to

conform with SBT instead of traditional MBT moved slowly (Wright, 2011). Wright (2009) asserted that reform needed in GA would not happen without industry leadership.

Reforms to GSC have moved slowly too. More explicit standards for ADM were included in the last revisions of several PTS; the *Instrument Rating PTS* and *Certified Flight Instructor, Instrument Rating PTS* were revised in 2010 (FAA, 2010a; FAA, 2010b). The *Private PTS* and *Commercial PTS* were revised to include similar, explicit standards for ADM and an emphasis on using SBT to evaluate applicants, effective June 2012 (FAA, 2011a; FAA, 2011b). The FAA also included extensive material on SBT, SA, and SRM in the latest version of the *Aviation Instructor's Handbook* (FAA, 2008a) and discussed ADM in the latest version of the *Pilot's Handbook of Aeronautical Knowledge* (FAA, 2008b).

### **Continuing the GA Training Reform Initiative**

GA safety and training were still a significant issue in 2012. The NTSB added GA Safety to its *Most Wanted List* in 2011 (NTSB, 2011c). The NTSB also hosted a forum, *General Aviation Safety: Climbing to the Next Level*, in June 2012 to raise awareness of GA safety issues, promote discussions, and determine effective actions to be taken (NTSB, 2012). Industry organizations such as AOPA, GAMA, and SAFE plus many individual researchers have expressed concerns over the sustainability of GA if accident rates are not improved (GAMA, 2010; SAFE, 2011b; Wright, 2009).

Continuing the GA training reform required active leadership from industry groups to promote the implementation of new training methods (Wright, 2011). SAFE is an industry group that has played an important role in coordinating the GA training reform since it formed in 2009. SAFE is an organization of aviation educators “fostering



professionalism and excellence in aviation through continuing education, professional standards, and accreditation” (“About SAFE,” 2012). SAFE’s Mission Statement states that they seek to “create a safer environment through enhanced education” (“Vision & Mission Statement,” 2012).

SAFE held a Pilot Training Reform Symposium in May 2011 where members met to form a “consistent framework for reform” (SAFE, 2011a). Attendees included senior FAA personnel, decorated flight instructors, and many prominent industry representatives (Stowell, 2011). SAFE published a preliminary report soon after the symposium, detailing the projects that SAFE members recommended. Members supplied many suggestions at the Symposium to promote GA training reform. These suggestions were consolidated into the following six “actionable and specific projects”:

1. Conduct a thorough general aviation fatal accident root cause analysis to pinpoint underlying accident causality as a means to create effective remedial actions.
2. Create a new flight review option that can be enabled as an FAA-sponsored pilot proficiency award program.
3. Revise FAA doctrine and standards to implement scenario-based testing, risk management, and other higher order pilot skills.
4. Modify flight instructor doctrine, initial testing, and renewal procedures to include the teaching of higher order pilot skills.
5. Implement voluntary flight instructor professional accreditation programs and continuing education that emphasize higher order pilot skills, scenario training, and interpersonal relationship skills.

6. Create and implement model curricula that incorporate higher order pilot skills, scenario-based training, and integration of simulation and other teaching methods to include interpersonal relationship skills. (SAFE, 2011a, p. 4, 10)

The report designated a project lead, participating organizations, required actions, expected outcomes, and a proposed timeline for each of the projects (SAFE, 2011a).

An update report, released in October of 2011, described the progress achieved since the Symposium (SAFE, 2011b). Limited progress had been made. For example, the FAA convened a Flight Training Standards Aviation Rulemaking Committee in late 2011 to specifically address needed reforms in standards for airmen knowledge tests and PTS (SAFE, 2011b). Several courseware providers posted free online training syllabi within weeks of the symposium (SAFE, 2011b). The FAA also revitalized and reorganized the General Aviation Joint Steering Committee (GAJSC), a committee composed of industry and FAA representatives that has existed but not provided much leadership for the past decade (SAFE, 2011b).

The update report also voiced concerns. Training reforms remained vital to improve aviation safety and to promote growth (SAFE, 2011b). SAFE called for leadership, particularly from the FAA and the GAJSC, and from the bottom up through grassroots organizations (SAFE, 2011b). The report called on the FAA and industry to reach a consensus on the path to training reform (SAFE, 2011b).

### **Summary**

The GA community incorporates a large and varied population, but has had a poor safety record compared to other aviation sectors. The number of active pilots, certificates issued, and hours flown all decreased in the last ten years, continuing the trend of the

preceding decades. Relevant subsets of the GA pilot population include novice pilots, accident-prone pilots, and ERAU student pilots.

The concepts of judgment, ADM, HOTS, and SA were researched. Government, academia, and industry formed the first formal ADM training efforts, employing behavior management strategies to improve ADM. ADM training later evolved to emphasize problem-solving skills and practice. Researchers created and tested different forms of PBL, including SBT, collaborative problem-solving, and case study. Researchers also tested the effectiveness of various memory aids and mnemonics.

ADM training became a cornerstone of the growing movement to reform GA training. The traditional MBT paradigm was no longer sufficient to improve ADM and HOTS in GA; training needed to shift to an SBT paradigm in order to effectively teach ADM and HOTS and improve GA's safety record. Significant efforts to reform GA training into an SBT paradigm included CGAR, the FITS program, and SAFE.

## **Chapter III**

### **Methodology**

#### **Research Approach**

This comparative study used the following experimental design. The population sample was divided into two groups – control and experimental – each with 15 participants. The control group completed two SBT sessions in a Frasca Mentor Advanced Aviation Training Device (AATD) (Frasca International, Inc., 2010). The initial AATD session was used to establish the participants' baseline ADM while the second AATD session revealed any change in the participants' ADM. Like the control group, the experimental group completed two SBT sessions in an AATD to establish a baseline and then document any change in the participants' ADM. The experimental group also received an ADM training treatment between the first and second SBT sessions. The researcher observed all of the AATD sessions, conducted the ADM training treatment, and scored the participants' ADM, based on real-time observations and subsequent review of the sessions via video and audio recordings.

#### **Design and procedures.**

*Designating the control and experimental groups.* The study participants were divided into two groups. The participants' names were entered into a computer program which randomly assigned a number to each participant. The participants were then organized according to his or her assigned number. Those participants with the lower 15 numbers were designated the control group; those with the higher 15 numbers were designated the experimental group.

*AATD session design.* The study used four ADM scenarios to test the participants' ADM in the AATD sessions. The four scenarios included a variety of different decisions that pilots realistically encounter on a visual flight rules (VFR) cross-country flight. The variety of scenarios allowed enough possible combinations of first and second scenarios to avoid rehearsal effects. The scenarios were randomly chosen for each participant's first AATD session. The scenario for that participant's second AATD session was randomly selected from the remaining three scenarios.

The researcher developed the four scenarios using personal experiences and stories from fellow flight instructors to generate points where the participant would need to make a decision. Personal knowledge of a typical ERAU flight student's training also informed several design choices for the scenarios (simulation equipment used, planned route of flight, etc.). The dilemmas presented in each scenario were generic and designed to imitate a common human-error related accident cause.

Each scenario simulated a VFR cross-country flight in a Cessna 172S NAV III Skyhawk (C172 Nav III) equipped with the Garmin G1000 avionics suite. The C172 Nav III is a single-engine propeller-driven aircraft with a 36-foot wingspan that can carry four people including the pilot. The C172 Nav III is the aircraft ERAU uses for all primary flight training. The Frasca Mentor AATD, installed at ERAU's Daytona Beach campus, is modeled after the C172 Nav III. Using that Frasca Mentor allowed the researcher to put the participants, who were all ERAU flight students, into a familiar aircraft, eliminating any effect on performance caused by an unfamiliar aircraft.

The route of flight departed from Southwest Florida International Airport (KRSW) near Naples, Florida and followed visual landmarks and Victor airways to

Hurlburt Field (KHRT) near Pensacola, FL. This route was chosen because it was far enough away from Daytona Beach to put students in unfamiliar territory but was still within the AATD's geographical database.

Each scenario began with the participant cruising in straight and level flight somewhere enroute. The starting location changed for each scenario. Otherwise, starting conditions were similar for all four scenarios. Appendix F shows the AATD Session Procedure for one of the scenarios which includes the initial conditions. While the scenario was in progress, a researcher assistant role-played as air traffic control and any other voices the participant would hear over the radio or in the cockpit.

*AATD session procedure.* Each participant completed two AATD sessions, spaced several days apart. The procedure for the first and second AATD sessions was identical. Prior to entering the AATD, the researcher briefed each participant on the schedule for the session. The participant was provided with a Consent Form (see Appendix B) and a Pilot Briefing (see Appendix C). The participant was then provided with cross-country planning materials including a weight and balance form, a standard weather briefing, a completed flight plan, a completed navigation log, and VFR sectional charts and IFR low-altitude enroute charts covering the entire route. The participants were allowed as much time as they desired to review and organize the cross-country materials.

When the participant was ready, the researcher guided them to the AATD and provided a notepad, pen, and an Airport Facility Directory (AFD). The researcher briefed the participant that he or she would start the simulation when ready by pushing a red button on the AATD's instrument panel. Participants were instructed to fly the AATD as

if they were flying a real Cessna 172 on a real VFR cross-country flight. The researcher stated that she would tell the participant to end the scenario by pushing the red button when it was time.

The researcher conducted a debrief at the end of each session, before the participant exited the AATD. The Debrief Form (see Appendix E) included several questions designed to assess the participant's SA and the degree to which he or she used an ADM process. The debrief was not treated as data but assisted the researcher in scoring the participant's ADM.

***Treatment design.*** The researcher designed the treatment, drawing guidance from the FAA's Advisory Circular concerning ADM, *Aeronautical Decision Making* (FAA, 1991), and from FITS materials including the *FAA-Industry Training Standards (FITS) Program Plan* (2003), *FITS Master Instructor Syllabus* (2006), and *SRM Scenario* (2009). The intent was to provide SBT to pilots in a classroom setting, thus accelerating the acquisition of higher-level decision-making skills. The treatment was divided into two phases: (a) inform the participants on ADM principles, strategies, and practical applications, and (b) guide the participants as they apply that information to a cross-country flight scenario.

The treatment was conducted in a conference-style classroom on ERAU's Daytona Beach campus. The room was equipped with a single, long table where everyone sat. This arrangement allowed the participants to see and interact with each other, encouraging everyone's involvement in the discussion.

***Treatment procedure.*** The treatment was delivered to the experimental group in groups of two to six participants between the participants' first and second AATD

sessions. Each treatment session began with providing a Consent Form to each participant (see Appendix B). The researcher then delivered a short presentation on ADM principles, strategies, and practical applications using a PowerPoint® presentation as a visual aid. The expected outcomes of the treatment were that participants:

- Accepted the importance of sound ADM
- Understood that ADM must be active; the pilot in command (PIC) must actively seek out decisions and then resolve them
- Understood and used the 5Ps (Plan, Plane, Pilot, Passengers, Programming), IMSAFE (Illness, Medication, Stress, Alcohol, Fatigue, Eating), DECIDE (Detect, Estimate, Choose, Identify, Do, Evaluate), and FATE (Fly the airplane, Assess the situation, Take action, Evaluate) models to analyze a scenario.

Next, the researcher guided the participants through a group discussion concerning a hypothetical cross-country flight. A PowerPoint presentation sourced from the FITS website (“*SRM Scenario*,” 2009) displayed relevant information about the flight while the researcher prompted the participants to apply the ADM information they had just received. Prompts included:

- What is the status of the 5Ps now?
- What are your concerns?
- Do you have any decisions to make? Explain.
- What actions could you take?
- What resources could you use to help make this decision?
- What action would you take? Why?



- What concerns do you have after you decided to ...?

The researcher ended the group session by recapping key points of the presentation, answering any lingering questions, and soliciting feedback on how useful the participants thought the session would be if it were integrated into ERAU's regular flight training curriculum.

**Apparatus and materials.** The AATD sessions were completed in a Frasca Mentor AATD (Frasca International, Inc., 2010) installed at ERAU's Daytona Beach campus (see Figure 1). This AATD's instrument panel included the G1000 integrated avionics suite, a popular example of an all-glass panel available in many GA TAAs. The G1000's two liquid crystal displays (LCDs) replace the traditional six-pack flight instruments and separate navigation and avionics components into a larger, integrated format. With a functional understanding of how to operate it, the G1000 can dramatically improve a pilot's SA by making flight information easier to scan and process ("*Garmin G1000*<sup>®</sup>," 2012). However, the G1000 can just as easily overwhelm a pilot who is unfamiliar with the system. Two video cameras installed on either side of the AATD and a portable audio recorder were used to record the sessions.

The treatment was conducted in a conference-style classroom on ERAU's Daytona Beach campus. The room was equipped with a single, long table. The researcher used a computer, a projector, and a hanging screen to display two PowerPoint presentations. The researcher used a whiteboard and markers to provide additional training material.



*Figure 1, Frasca Mentor Advanced Aviation Training Device (AATD).*

### **Population Sample**

The population was collegiate flight students. Participants for this study were solicited by the researcher from the population of flight students at ERAU's Daytona Beach campus. The researcher visited all fall 2011 sections of AS 321 Commercial Pilot Operations (ERAU's commercial pilot ground lab) and one of the weekly meetings of Alpha Omicron Alpha, an aviation honor society. These two population subsets included approximately 170 students.

The researcher gave a short presentation describing the study and its benefits, and then asked those interested to provide their contact information. Seventy students volunteered contact information. Many students stopped responding to emails, but 32 remained in contact. The study thus began with a population sample of 32 active

participants. Two of those participants did not complete the last phase of the study, leaving the study with a final self-selected sample of 30 participants.

### **Data Collection Device**

This study used one primary data collection device. The researcher used the Scoring Sheet (see Appendix D) to record observations of the participants and assign scores as they completed each AATD session. Scores were entered for each of the variables at each decision point in the session. Each ordinal variable was scored as a 1, 2, or 3. A score of 2 described a relatively wide range of behaviors and meant that the researcher judged the participant's ADM to be adequate for the situation but not exceptional. A score of 1 meant that the researcher judged the participant's ADM to be inadequate for the situation. A score of 3 meant that the researcher judged the participant's ADM to be exceptionally good, not merely good enough. Later, the scores for each ordinal variable (Problem Comprehended, Projection, Decision Process Used, and Timely Manner) were averaged across all decision points to yield a session score for each variable. The session scores were then averaged to generate an Overall ADM score for that participant.

For example, consider the scores received by Participant 1 in Appendix D, Sample Scoring Sheet. This sample shows scores that four generic participants could have received in the first AATD session. Participant 1 completed a session that involved three decision points. In Comprehension, Participant 1 received a score of 2 for the first decision point, a 1 for the second decision point, and a 2 for the third decision point. The researcher calculated that participant's session score for Comprehension by averaging those three scores (2, 1, and 2) which results in a score of 1.67. The researcher averaged

the remaining ordinal scores to calculate a session score for Projection (1.67), Decision Process Used (1.67), and Timely Manner (2.00). The researcher then averaged the session scores (1.67, 1.67, 1.67, and 2.00) to calculate an Overall ADM score of 1.75. Nominal data (Problem Detected, Problem Resolved, and Safe Outcome) was aggregated and analyzed by group – control versus experimental.

The researcher also used a Debrief Form (see Appendix E) to facilitate a guided debrief with the participant at the end of each AATD session. The debrief served two purposes: (a) the debrief was intended to maximize participants' learning by guiding them through a review of the experiences and the decisions they just made, and (b) the debrief solicited the participants' thoughts and impressions which provided valuable insight for the researcher and enabled more accurate scoring.

**Content validity and reliability.** Content validity is an estimate of how well an instrument reflects the intended construct or domain of content (Howell et al., 2005). In this study, the researcher sought to assess ADM and designed a study to test for ADM skills. The process for determining content validity involved using experiential content validity experts (CFIs) and professional experts (aviation practitioners who were university professors). The content validity of the study was supported by the theory-based constructs from the literature review on ADM. The result was a study design that had content validity with greater relevance for the target population of pilots ranging from student pilots to apprentice flight instructors.

Content reliability refers to whether an instrument will produce consistent results each time it is administered in the same setting to the same subject (George & Mallery, 2011). The researcher determined content reliability by performing a series of

Chronbach's alpha analyses, which measure an instrument's internal consistency. The Chronbach's alpha was measured for four subsets of data. The first measurement considered the ordinal variable scores for all participants in Round One, excluding their Overall ADM scores. The second analysis measured the consistency of the ordinal variable scores for all participants in Round Two, excluding their Overall ADM scores. The third analysis measured the consistency of all ordinal variable scores for all participants in Round One, including the Overall ADM scores; the last analysis measured the consistency of all ordinal variable scores for all participants in Round Two.

### **Treatment of the Data**

**Descriptive statistics.** The data collection device for this study, the Scoring Sheet, recorded the following variables: Problem Detected, Problem Comprehended, Projection, Decision Process Used, Problem Resolved, Timely Manner, and Safe Outcome (see Appendix D). Three variables (Problem Detected, Problem Resolved, and Safe Outcome) were nominal and were described in charts. Four variables (Problem Comprehended, Projection, Decision Process Used, and Timely Manner) were ordinal data and were described in tables depicting the mean, the standard deviation (SD), the minimum, the maximum, and the count (N).

**Reliability Testing.** The researcher determined content reliability by performing a series of Chronbach's alpha analyses. Chronbach's alpha measures were performed including all of the ordinal values except for the Overall ADM scores. Then, Chronbach's alpha measures were performed again for all of the ordinal values including the Overall ADM scores.

**Hypothesis testing.** Hypothesis testing was conducted in five stages. First, the participants' baseline ADM performance was established by using Mann-Whitney tests to compare the scores for the control and the experimental groups in the first AATD session. Second, the participants' ending ADM was tested by using Mann-Whitney tests to compare the scores for the control and the experimental groups in the second AATD session. Third, the change in the control group's ADM was tested by using Wilcoxon tests to compare the control group's scores from the first AATD sessions to their scores in the second AATD sessions. Fourth, the change in the experimental group's ADM was tested by using Wilcoxon tests to compare the experimental group's scores from the first AATD sessions to their scores from the second AATD sessions. Fifth, Mann-Whitney tests were used to compare the Delta, or change, in ADM for the control group to the Delta in ADM for the experimental group.

## Chapter IV

### Results

#### Descriptive Statistics

The sample included 30 flight students enrolled in a baccalaureate program at ERAU. The sample included 23 male students and 7 female students. The control group was composed of 12 male students and 3 female students; the experimental group was composed of 11 male students and 4 female students.

All participants had relatively low total flight time. Total time was recorded from each participant's logbook at the beginning of his or her first AATD session. The minimum total time was 55 hours and the maximum total time was 510 hours. Twenty-three of the participants (out of 30 total) had fewer than 200 hours. Figure 2 shows the range of total time of all the participants, and Figure 3 shows the range of total time of the participants for the control and the experimental groups (grouped in 50-hour intervals).

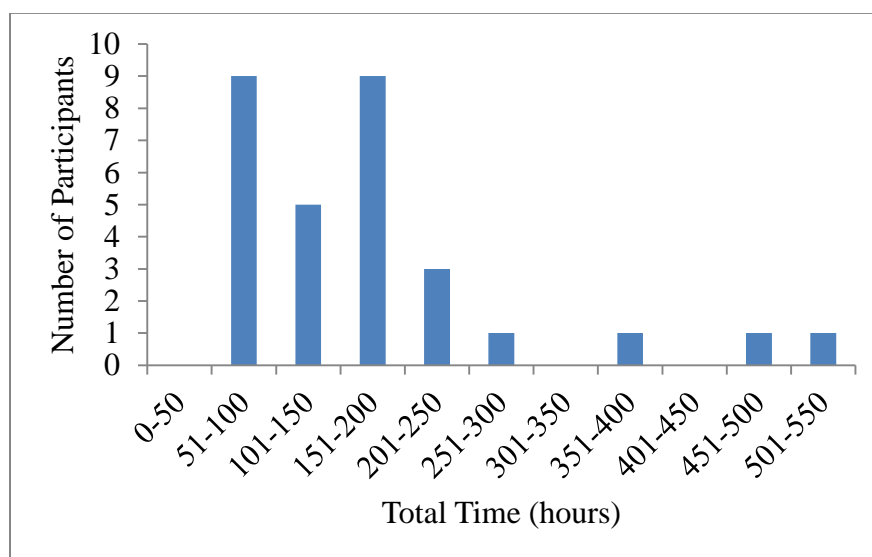


Figure 2. Total time of all participants.

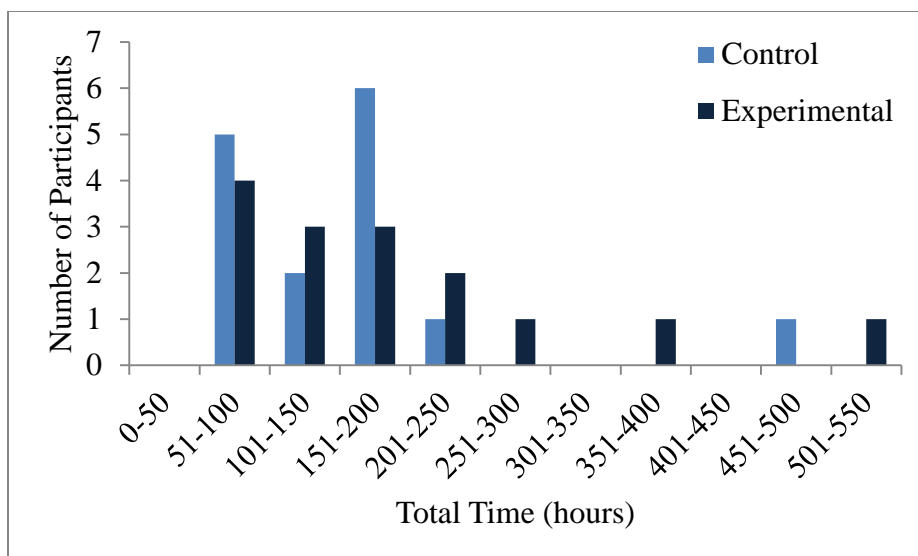


Figure 3. Total time of the participants, control and experimental groups.

The participants held a varying range of pilot certificates and ratings. Figure 4 shows a count of participants by the most advanced rating held. Figure 5 shows a count of participants by the most advanced rating held for the control and the experimental groups. The mode of the sample was the private pilot certificate. The population also included nine commercial pilots and four student pilots. Four of the commercial pilots were also CFIs. One participant who held a glider rating was categorized according to that participant's highest airplane certificate, private pilot certificate with multi-engine rating.



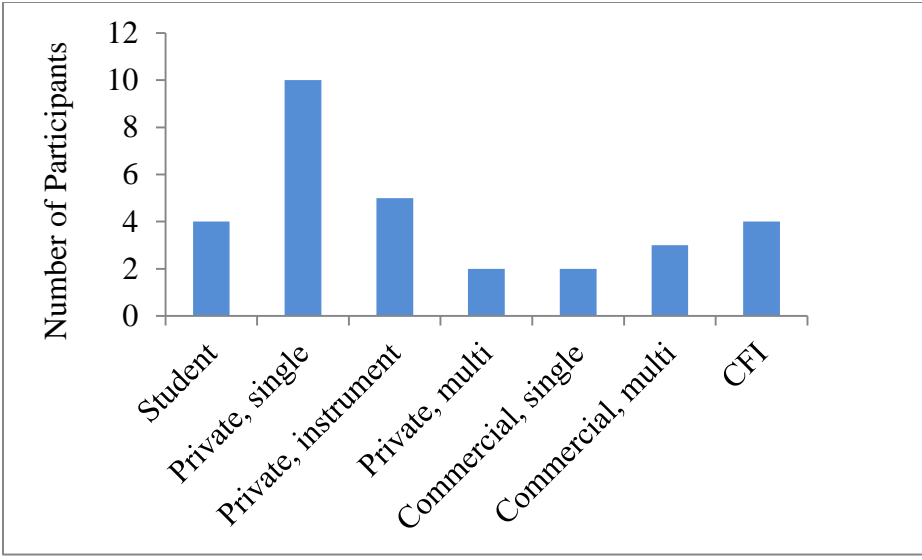


Figure 4. Highest certification held by the participants.

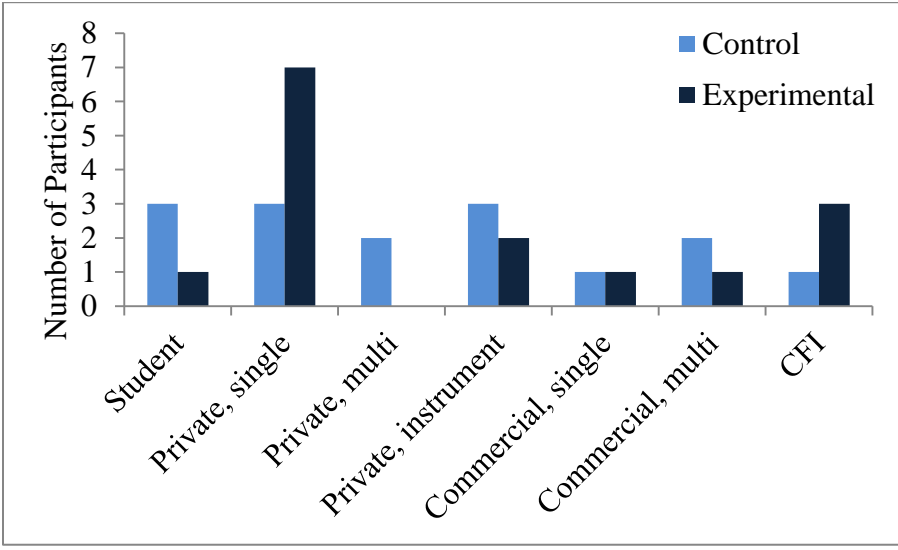


Figure 5. Highest certification held by the participants, control and experimental groups.

**Dependent variables.** Each of the participants was scored at multiple decisions points within each AATD session using seven variables. The data were either nominal or ordinal as follows:

- Nominal Variables
  - Problem Detected (Yes/No)
  - Problem Resolved (Yes/No)
  - Safe Outcome (Yes/No)
- Ordinal Variables
  - Problem Comprehended (1, 2, or 3)
  - Problem Projected (1, 2, or 3)
  - Decision Process Used (1, 2, or 3)
  - Timely manner (1, 2, or 3)

The researcher then averaged the session scores to generate an overall ADM score for that participant in that AATD session. Finally, the researcher calculated the Delta (change) in each participant's ADM between Round One and Round Two. These calculations yielded the following ordinal variables:

- Ordinal Variables, Calculated
  - Overall ADM (0 - 3)
  - Delta Problem Comprehended (0 - 3)
  - Delta Problem Projected (0 - 3)
  - Delta Decision Process Used (0 - 3)
  - Delta Timely manner (0 - 3)
  - Delta Overall ADM (0 - 3)

The researcher counted the nominal data and aggregated them for the two groups, control or experimental. Figures 6 through 8 show the percentage of decisions for which the participants received a *Yes* score instead of a *No*. For Problem Detected (Figure 6), a

Yes means the participant detected the problem. For Problem Resolved (Figure 7), a Yes means the participant successfully resolved the problem. For Safe Outcome (Figure 8), a Yes means that the participant overcame the problem to reach a safe outcome.

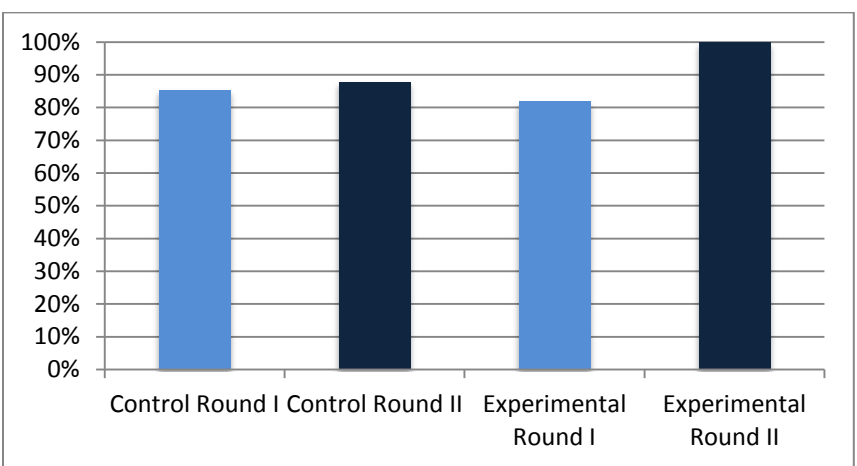


Figure 6. Participants detected the problem, percentage of decisions.

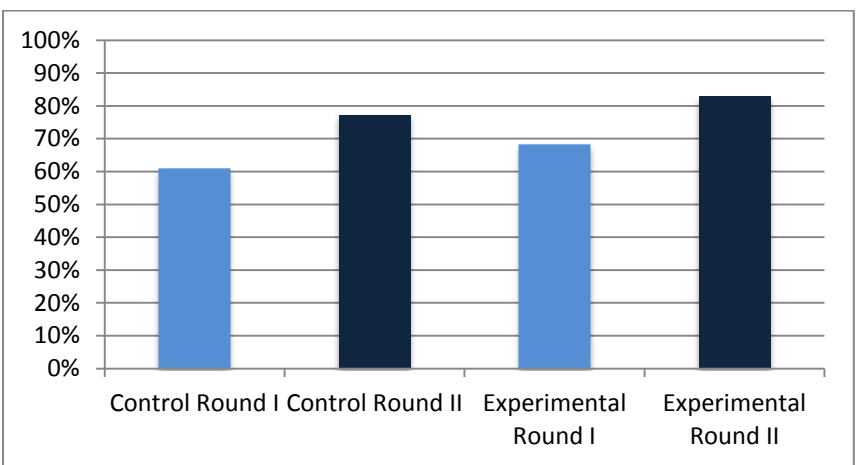
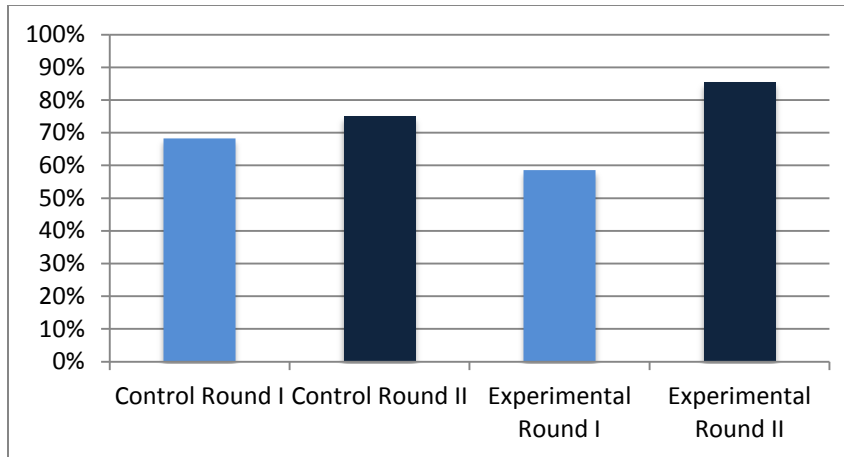


Figure 7. Participants resolved the problem, percentage of decisions.



*Figure 8.* Participants reached a safe outcome.

Table 1 shows the five ordinal variables (Problem Comprehended, Problem Projected, Decision Process Used, Timely Manner, and Overall ADM) for the control and experimental groups in the first AATD sessions. Very few participants received 3s. Slightly more received 1s, and the majority received 2s.

Table 1

*Ordinal Variables for the Control and Experimental Groups, First AATD Sessions*

		Comp	Proj	Proc	Time	Overall
Control	N	15	15	15	15	15
	Mean	1.78	1.76	1.64	1.60	1.69
	Range	1.33	1.67	1.33	1.50	1.33
	Minimum	1.00	1.00	1.00	1.00	1.00
	Maximum	2.33	2.67	2.33	2.50	2.33
	Std. Deviation	.32	.42	.44	.54	.40
Experimental	N	15	15	15	15	15
	Mean	1.88	1.82	1.74	1.82	1.82
	Range	1.67	2.00	2.00	2.00	1.92
	Minimum	1.33	1.00	1.00	1.00	1.08
	Maximum	3.00	3.00	3.00	3.00	3.00
	Std. Deviation	.47	.65	.71	.74	.63

*Note.* Comp = Problem Comprehended, Proj = Problem Projected, Proc = Decision Process Used, Time = Timely Manner, Overall = Overall ADM.

Table 2 shows the five ordinal variables for the control and experimental groups in the second AATD sessions.

Table 2

*Ordinal Variables for the Control and Experimental Groups, Second AATD Sessions*

		Comp	Proj	Proc	Time	Overall
Control	N	15	15	15	15	15
	Mean	2.02	1.96	1.83	1.82	1.91
	Range	2.00	2.00	2.00	2.00	2.00
	Minimum	1.00	1.00	1.00	1.00	1.00
	Maximum	3.00	3.00	3.00	3.00	3.00
	Std. Deviation	.58	.68	.69	.74	.65
Experimental	N	15	15	15	15	15
	Mean	2.02	1.93	1.89	1.96	1.95
	Range	.67	1.67	1.33	1.67	1.17
	Minimum	1.67	1.00	1.00	1.00	1.25
	Maximum	2.33	2.67	2.33	2.67	2.42
	Std. Deviation	.20	.42	.37	.45	.33

*Note.* Comp = Problem Comprehended, Proj = Problem Projected, Proc = Decision Process Used, Time = Timely Manner, Overall = Overall ADM.

Table 3 shows the Delta, or change, in the five ordinal variables between the first and second AATD sessions for the control and experimental groups. The Delta was calculated by subtracting each participant's Round One scores from their respective

Round Two scores. A positive Delta indicated an improvement in ADM; a negative Delta indicated a regression.

Table 3

*Delta of Ordinal Variables for the Control and Experimental Groups*

		Delta Comp	Delta Proj	Delta Proc	Delta Time	Delta Overall
Control	N	15	15	15	15	15
	Mean	.24	.20	.19	.22	.21
	Range	1.33	1.33	1.50	2.00	1.21
	Minimum	-.33	-.33	-.50	-.67	-.38
	Maximum	1.00	1.00	1.00	1.33	.83
	Std. Deviation	.42	.43	.50	.63	.44
Experimental	N	15	15	15	15	15
	Mean	.14	.11	.14	.13	.13
	Range	1.67	2.00	2.33	2.67	2.00
	Minimum	-1.00	-1.33	-1.33	-1.33	-1.25
	Maximum	.67	.67	1.00	1.33	.75
	Std. Deviation	.40	.56	.72	.69	.56

*Note.* Comp = Problem Comprehended, Proj = Problem Projected, Proc = Decision Process Used, Time = Timely Manner, Overall = Overall ADM.

### Reliability Testing

A series of Chronbach's alpha analyses was conducted to measure the reliability of the data. The first analysis measured the consistency of the ordinal variable scores for

all participants in Round One, excluding their Overall ADM scores. The Chronbach's alpha on standardized items for those four variables was 0.966. The second analysis measured the consistency of the ordinal variable scores for all participants in Round Two, excluding their Overall ADM scores. The Chronbach's alpha on standardized items for those four variables was 0.963.

The next analyses included the scores for Overall ADM. The third analysis measured the consistency of all ordinal variable scores for all participants in Round One. The Chronbach's alpha on standardized items for those five variables was 0.980. The fourth analysis measured the consistency of all ordinal variable scores for all participants in Round Two. The Chronbach's alpha on standardized items for those five variables was 0.978.

### **Hypothesis Testing**

Several Mann-Whitney tests and Wilcoxon tests were calculated to test the null hypothesis - There was no difference in demonstrated ADM between pilots who received specialized ADM training and pilots who received no specialized training, for flight students enrolled in a baccalaureate program at ERAU.

**Baseline ADM performance.** A Mann-Whitney test was calculated to test the null hypothesis - There was no difference in Round One Problem Comprehended between the control and experimental groups. Figure 9 shows the results. There was no difference in Round One Problem Comprehended between the control and experimental groups.



### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Round_ONE_Comprehended is the same across categories of Control_Experimental.	Independent-Samples Mann-Whitney U Test	.828	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 9. Round One Problem Comprehended scores, Control and Experimental groups.

A Mann-Whitney test was calculated to test the null hypothesis - There was no difference in Round One Problem Projected between the control and experimental groups. Figure 10 shows the results. There was no difference in Round One Problem Projected between the control and experimental groups.

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Round_ONE_Projection is the same across categories of Control_Experimental.	Independent-Samples Mann-Whitney U Test	.883	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 10. Round One Problem Projected scores, Control and Experimental groups.

A Mann-Whitney test was calculated to test the null hypothesis - There was no difference in Round One Decision Process Used between the control and experimental groups. Figure 11 shows the results. There was no difference in Round One Decision Process Used between the control and experimental groups.

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Round_ONE_Decision_Process_Used is the same across categories of Control_Experimental.	Independent-Samples Mann-Whitney U Test	.899	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 11. Round One Decision Process Used scores, Control and Experimental groups.

A Mann-Whitney test was calculated to test the null hypothesis - There was no difference in Round One Timely Manner between the control and experimental groups. Figure 12 shows the results. There was no difference in Round One Timely Manner between the control and experimental groups.

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Round_ONE_Timely_Manner is the same across categories of Control_Experimental.	Independent-Samples Mann-Whitney U Test	.485	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 12. Round One Timely Manner scores, Control and Experimental groups.

A Mann-Whitney test was calculated to test the null hypothesis - There was no difference in Round One Overall ADM between the control and experimental groups. Figure 13 shows the results. There was no difference in Round One Overall ADM between the control and experimental groups.

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Round_ONE_Overall_ADM is the same across categories of Control_Experimental.	Independent-Samples Mann-Whitney U Test	.787	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 13. Round One Overall ADM scores, Control and Experimental groups.

**Ending ADM performance.** A Mann-Whitney test was calculated to test the null hypothesis - There was no difference in Round Two Problem Comprehended between the control and experimental groups. Figure 14 shows the results. There was no difference in Round Two Problem Comprehended between the control and experimental groups.

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Round_TWO_Comprehended is the same across categories of Control_Experimental.	Independent-Samples Mann-Whitney U Test	.476	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 14. Round Two Problem Comprehended scores, Control and Experimental groups.

A Mann-Whitney test was calculated to test the null hypothesis - There was no difference in Round Two Problem Projected between the control and experimental groups. Figure 15 shows the results. There was no difference in Round Two Problem Projected between the control and experimental groups.

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Round_TWO_Projection is the same across categories of Control_Experimental.	Independent-Samples Mann-Whitney U Test	.750	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 15. Round Two Problem Projected scores, Control and Experimental groups.

A Mann-Whitney test was calculated to test the null hypothesis - There was no difference in Round Two Decision Process Used between the control and experimental groups. Figure 16 shows the results. There was no difference in Round Two Decision Process Used between the control and experimental groups.

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Round_TWO_Decision_process_used is the same across categories of Control_Experimental.	Independent-Samples Mann-Whitney U Test	.327	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 16. Round Two Decision Process Used scores, Control and Experimental groups.

A Mann-Whitney test was calculated to test the null hypothesis - There was no difference in Round Two Timely Manner between the control and experimental groups. Figure 17 shows the results. There was no difference in Round Two Timely Manner between the control and experimental groups.

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Round_TWO_Timely_Manner is the same across categories of Control_Experimental.	Independent-Samples Mann-Whitney U Test	.409	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 17. Round Two Timely Manner scores, Control and Experimental groups.

A Mann-Whitney test was calculated to test the null hypothesis - There was no difference in Round Two Overall ADM between the control and experimental groups. Figure 18 shows the results. There was no difference in Round Two Overall ADM between the control and experimental groups.

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Round_TWO_Overall_ADM is the same across categories of Control_Experimental.	Independent-Samples Mann-Whitney U Test	.307	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 18. Round Two Overall ADM scores, Control and Experimental groups.

**Change in ADM for the control group.** A Wilcoxon test was calculated to test the null hypothesis - There was no difference in Problem Comprehended between Round One and Round Two for the control group. Figure 19 shows the results. There was a difference in Problem Comprehended between Round One and Round Two for the control group.

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between Round_ONE_Comprehended and Round_TWO_Comprehended equals 0.	Related-Samples Wilcoxon Signed Rank Test	.037	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

*Figure 19.* Problem Comprehended scores in Round One and Round Two for the Control group.

A Wilcoxon test was calculated to test the null hypothesis - There was no difference in Problem Projected between Round One and Round Two for the control group. Figure 20 shows the results. There was no difference in Problem Projection between Round One and Round Two for the control group.

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between Round_ONE_Projection and Round_TWO_Projection equals 0.	Related-Samples Wilcoxon Signed Rank Test	.085	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

*Figure 20.* Problem Projected scores in Round One and Round Two for the Control group.

A Wilcoxon test was calculated to test the null hypothesis - There was no difference in Decision Process Used between Round One and Round Two for the control group. Figure 21 shows the results. There was no difference in Decision Process Used between Round One and Round Two for the control group.

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between Round_ONE_Decision_Process_Used and Round_TWO_Decision_process_used equals 0.	Related-Samples Wilcoxon Signed Rank Test	.145	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

*Figure 21.* Decision Process Used scores in Round One and Round Two for the Control group.

A Wilcoxon test was calculated to test the null hypothesis - There was no difference in Timely Manner between Round One and Round Two for the control group. Figure 22 shows the results. There was no difference in Timely Manner between Round One and Round Two for the control group.

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between Round_ONE_Timely_Manner and Round_TWO_Timely_Manner equals 0.	Related-Samples Wilcoxon Signed Rank Test	.134	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

*Figure 22.* Timely Manner scores in Round One and Round Two for the Control group.

A Wilcoxon test was calculated to test the null hypothesis - There was no difference in Overall ADM between Round One and Round Two for the control group. Figure 23 shows the results. There was no difference in Overall ADM between Round One and Round Two for the control group.

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between Round_ONE_Overall_ADM and Round_TWO_Overall_ADM equals 0.	Related-Samples Wilcoxon Signed Rank Test	.089	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 23. Overall ADM scores in Round One and Round Two for the Control group.

**Change in ADM for the experimental group.** A Wilcoxon test was calculated to test the null hypothesis - There was no difference in in Problem Comprehended between Round One and Round Two for the experimental group. Figure 24 shows the results. There was no difference in Problem Comprehended between Round One and Round Two for the experimental group.

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between Round_ONE_Comprehended and Round_TWO_Comprehended equals 0.	Related-Samples Wilcoxon Signed Rank Test	.104	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 24. Problem Comprehended scores in Round One and Round Two for the Experimental group.

A Wilcoxon test was calculated to test the null hypothesis - There was no difference in Problem Projected between Round One and Round Two for the experimental group. Figure 25 shows the results. There was no difference in Problem Projected between Round One and Round Two for the experimental group.



### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between Round_ONE_Projection and Round_TWO_Projection equals 0.	Related-Samples Wilcoxon Signed Rank Test	.238	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

*Figure 25.* Problem Projected scores in Round One and Round Two for the Experimental group.

A Wilcoxon test was calculated to test the null hypothesis - There was no difference in Decision Process Used between Round One and Round Two for the experimental group. Figure 26 shows the results. There was no difference in Decision Process Used between Round One and Round Two for the experimental group.

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between Round_ONE_Decision_Process_Used and Round_TWO_Decision_process_used equals 0.	Related-Samples Wilcoxon Signed Rank Test	.452	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

*Figure 26.* Decision Process Used scores in Round One and Round Two for the Experimental group.

A Wilcoxon test was calculated to test the null hypothesis - There was no difference in Timely Manner between Round One and Round Two for the experimental group. Figure 27 shows the results. There was no difference in Timely Manner between Round One and Round Two for the experimental group.

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between Round_ONE_Timely_Manner and Round_TWO_Timely_Manner equals 0.	Related-Samples Wilcoxon Signed Rank Test	.382	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

*Figure 27.* Timely Manner scores in Round One and Round Two for the Experimental group.

A Wilcoxon test was calculated to test the null hypothesis - There was no difference in Overall ADM between Round One and Round Two for the experimental group. Figure 28 shows the results. There was no difference in Overall ADM between Round One and Round Two for the experimental group.

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between Round_ONE_Overall_ADM and Round_TWO_Overall_ADM equals 0.	Related-Samples Wilcoxon Signed Rank Test	.232	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

*Figure 28.* Overall ADM scores in Round One and Round Two for the Experimental group.

**Comparing delta between control and experimental groups.** A Mann-Whitney test was calculated to test the null hypothesis - There was no difference in Delta Problem Comprehended (the change between Round One and Round Two) between the control and experimental groups. Figure 29 shows the results. There was no difference in Delta Problem Comprehended between the control and the experimental groups.

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Delta_Comprehension is the same across categories of Control_Experimental.	Independent-Samples Mann-Whitney U Test	.847	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 29. Delta Problem Comprehended for the Control and Experimental groups.

A Mann-Whitney test was calculated to test the null hypothesis - There was no difference in Delta Problem Projected between the control and experimental groups.

Figure 30 shows the results. There was no difference in Delta Problem Projected between the control and the experimental groups.

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Delta_Projection is the same across categories of Control_Experimental.	Independent-Samples Mann-Whitney U Test	.899	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 30. Delta Problem Projected for the Control and Experimental groups.

A Mann-Whitney test was calculated to test the null hypothesis - There was no difference in Delta Decision Process Used between the control and experimental groups.

Figure 31 shows the results. There was no difference in Delta Decision Process Used between the control and the experimental groups.

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Delta_Decision_Process_Used is the same across categories of Control_Experimental.	Independent-Samples Mann-Whitney U Test	.916	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 31. Delta Decision Process Used for the Control and Experimental groups.

A Mann-Whitney test was calculated to test the null hypothesis - There was no difference in Delta Timely Manner between the control and experimental groups. Figure 32 shows the results. There was no difference in Delta Timely Manner between the control and the experimental groups.

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Delta_Timely_Manner is the same across categories of Control_Experimental.	Independent-Samples Mann-Whitney U Test	.801	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 32. Delta Timely Manner for the Control and Experimental groups.

A Mann-Whitney test was calculated to test the null hypothesis - There was no difference in Delta Overall ADM between the control and experimental groups. Figure 33 shows the results. There was no difference in Delta Overall ADM between the control and the experimental groups.

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Delta_Overall_ADM is the same across categories of Control_Experimental.	Independent-Samples Mann-Whitney U Test	.852	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

*Figure 33.* Delta Overall ADM for the Control and Experimental groups.

## Chapter V

### Discussion, Conclusions, and Recommendations

#### Discussions

**Significance of results.** The data from this study did not produce many statistically significant results or provide overwhelming support for the research hypothesis. Several factors could have contributed to these results.

**Experiment design factors.**

*Impact of variations in the sample.* The population for the experiment was more varied than originally intended by the researcher; thus, subsets of the population were small. The variation in ratings and certificates yielded interesting results, both expected and unexpected. Too few participants fell into the separate categories to make any generalizations, but the variation between how participants with different certificates fared suggest further research is required. More research questions include: (a) when is the best time to introduce ADM training? and (b) how effective are the current MBT curricula, compared to SBT curricula for teaching ADM?

Several comparisons can be made based on the researcher's observations. Table 4 compares the Overall ADM scores and Deltas for the participants. Participants are organized by whether they were part of the control or the experimental group and by highest certification (CFIs were also commercial pilots but were separated into their own category). Table 5 compares the same data for the participants but organizes them into CFI and non-CFI participants.

Table 4

*Overall ADM and Delta Overall ADM for Participants Grouped By Highest Certification: Control, Experimental, and All*

	Count	Round I	Round II	Delta
<b>Student</b>				
Control	3	1.21	1.39	0.18
Experimental	1	1.13	1.25	0.13
All Student	4	1.19	1.35	0.17
<b>Private</b>				
Control	8	1.80	2.03	0.22
Experimental	9	1.64	1.97	0.33
All Private	17	1.72	2.00	0.28
<b>Commercial</b>				
Control	3	1.75	2.17	0.42
Experimental	2	2.29	2.17	-0.13
All Commercial	5	1.97	2.17	0.20
<b>CFI</b>				
Control	1	2.13	1.75	-0.38
Experimental	3	2.25	1.97	-0.28
All CFI	4	2.22	1.92	-0.30

Regardless of whether they were given the experimental SBT ADM training, the student pilots generally scored lower than the higher-certificated participants.

Anecdotally (the number of student pilots is small), the three control student pilot participants averaged higher ADM scores in both rounds and a higher Delta than the lone experimental student pilot participant. This suggests that the training provided could be modified to better address student pilots' level of experience.

The control private pilot participants averaged a high Delta of 0.22, suggesting that private pilots benefit from ADM training whether it is MBT or SBT. The

experimental private pilots averaged the highest Delta Overall ADM of the experimental participants (0.33). The private pilots' improvement supports other research speculating that newly certificated private pilots would benefit greatly from ADM training (Adams, Hamilton, Koonce, & Hwoschinsky, 2002). That the experimental private pilot participants showed such great improvement suggests the experimental treatment was particularly effective for these participants. The experimental private pilots' higher Delta suggests that SBT ADM training was more effective than the MBT received by the control private pilot participants.

The three control commercial pilot participants averaged the highest Delta Overall ADM score of the control group (0.42). The two experimental commercial pilot participants averaged the highest Overall ADM score in Round One and tied with the three Round Two control commercial pilot participants for the highest Round Two Overall ADM scores. The experimental commercial pilot participants' Round Two scores were slightly lower than their Round One scores, though, resulting in a negative Delta Overall ADM. These results are inconclusive on whether SBT ADM training benefits commercial pilots more than MBT.

All of the CFI participants outperformed the other participants in the first AATD sessions. However, the CFI participants regressed in the second AATD sessions leading to negative Deltas (-0.38 and -0.28 for the control and experimental groups respectively). Working as a CFI adds the task of providing instruction to the mental workload that a GA pilot would otherwise be responsible for. The researcher therefore expected the CFIs to exhibit more complete SA and more efficient ADM, as they did in the first AATD sessions. Although the researcher expected that there would be less room for



improvement for CFIs than for non-CFIs, the researcher still expected a positive change in ADM. There was no clear reason why the CFIs regressed as a group in the second AATD sessions, other than the small number of CFIs ( $N=4$ ).

Table 5

*Overall ADM and Delta Overall ADM for CFI and Non-CFI Participants: Control, Experimental, and All*

	Count	Round I	Round II	Delta
CFI				
Control	1	2.13	1.75	-0.38
Experimental	3	2.25	1.97	-0.28
All CFI	4	2.22	1.92	-0.30
Non-CFI				
Control	14	1.66	1.92	0.26
Experimental	12	1.71	1.94	0.24
All Non-CFI	26	1.68	1.93	0.25

Figures 34, 35, and 36 show the distribution of participants' Overall ADM scores in a linear regression for the first and second AATD sessions, respectively, compared to their total time. ADM literature led the researcher to expect that a participant's experience would increase with total time, leading to a correlation between a participant's ADM scores and his or her total time. The researcher also expected the SBT ADM training to improve the participants' ADM above what they would have demonstrated without the training.

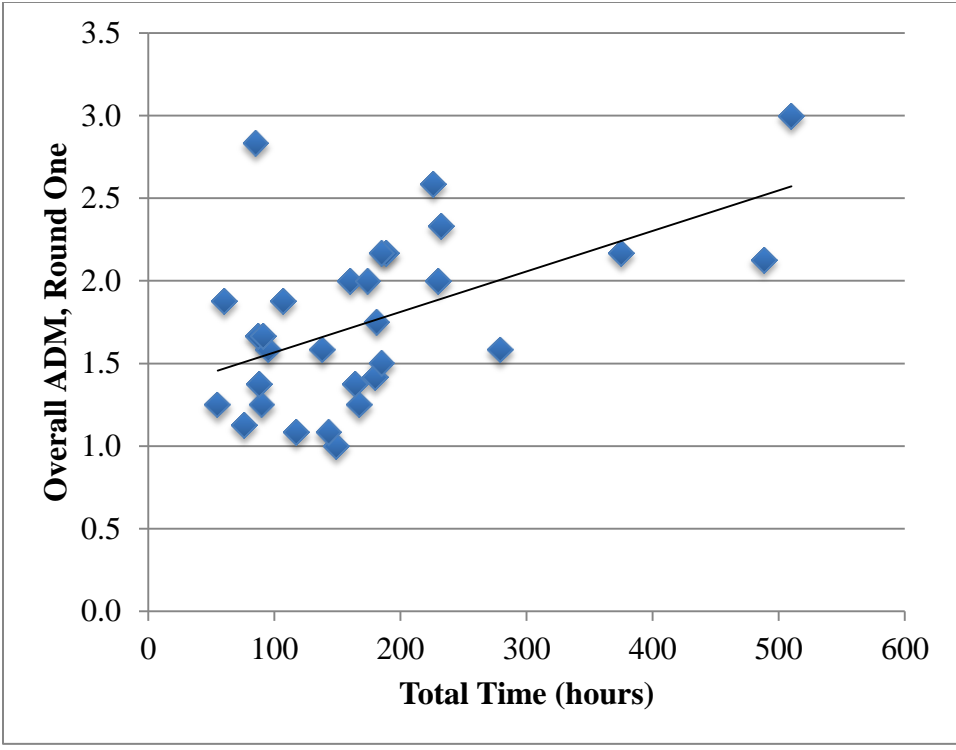


Figure 34. Total time and Overall ADM for all participants in Round One

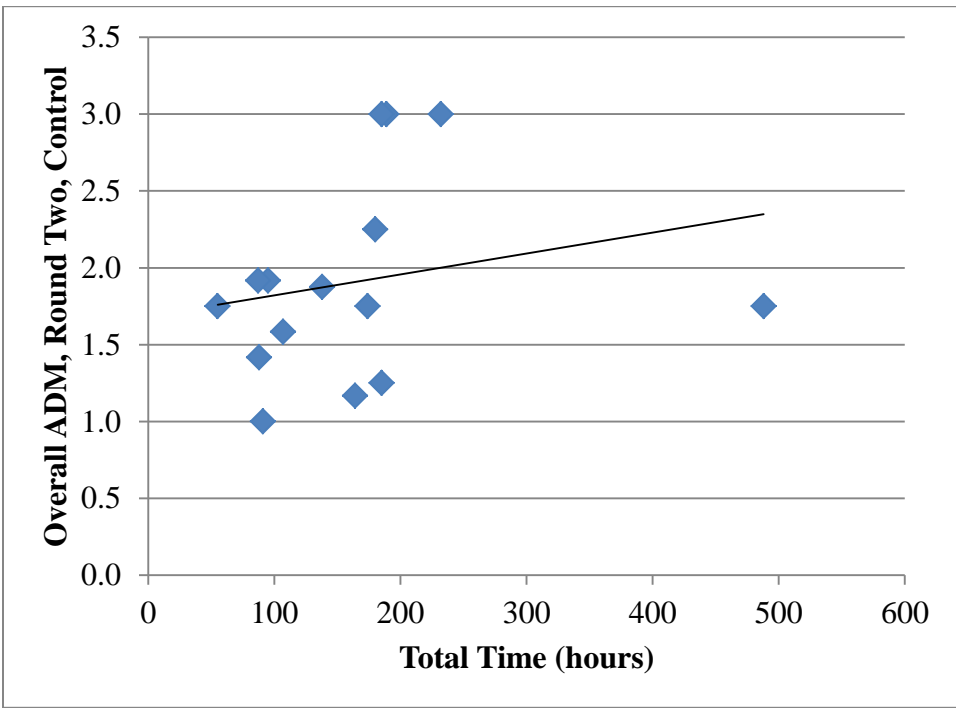


Figure 35. Total time and Overall ADM for the control participants in Round Two

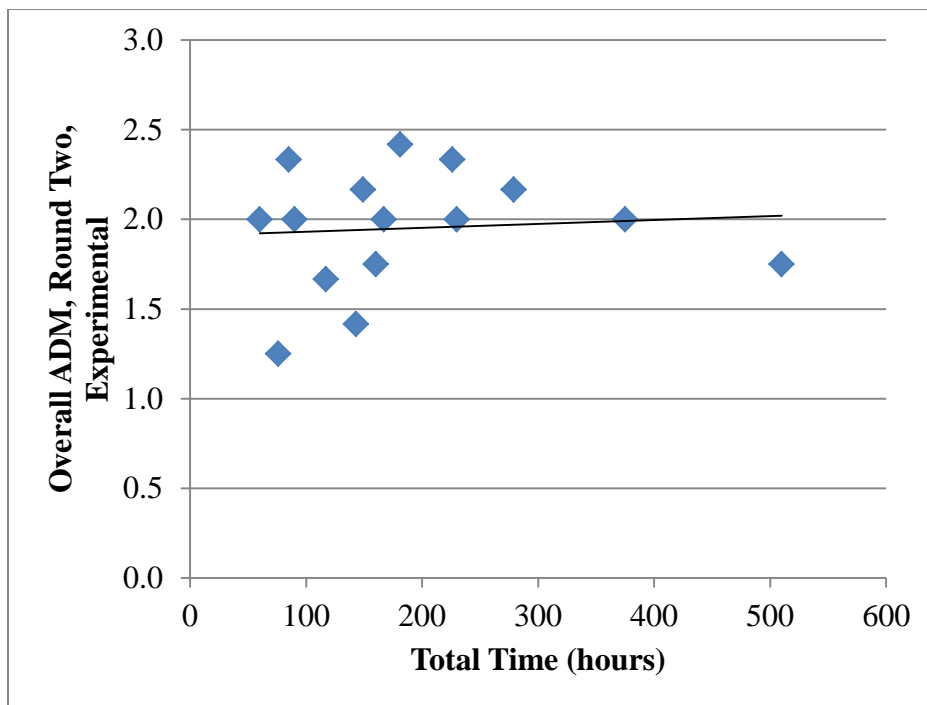


Figure 36. Total time and Overall ADM for the experimental participants in Round Two

An examination of Figure 34 shows an expected positive correlation between total time and overall ADM scores in the first AATD sessions. The positive correlation between total time and overall ADM scores is also present in the second AATD sessions as seen in Figures 35 and 36. However, a comparison of the two linear regressions in Figures 35 and 36 with the regression in Figure 34 shows that the positive correlation between total time and overall ADM is weaker for the control participants in the second AATD session and almost non-existent for the experimental participants in the second AATD session. The weakened correlation between increasing total time and improved ADM for the experimental participants in Round Two suggests that the experimental SBT ADM training was effective at developing lower time pilots' ADM.

Comparisons here are anecdotal since the sample for each experience level was not sufficient to allow meaningful statistical analysis. Further research is needed to determine whether these conclusions can be generalized.

***G1000 proficiency.*** Participants' success could have been influenced by their proficiency with the G1000. All students had some experience with it, but the skill level ranged from completely ignoring the system, through trying to use the very high frequency (VHF) omnidirectional radio ranges (VORs) but not the GPS moving map, through programming the entire flight plan into the MFD. Those who did not use the moving map generally had a difficult time locating themselves at the beginning of the scenario and did poorly when confronted with a significant decision later in the scenario. Others persisted in inept attempts to use a particular feature of the G1000, resulting in a higher workload and poorer ADM. Both IFR and VFR sectional charts were available to the participants, and the AATD was limited to a single visual display. Participants could choose whether to primarily use VFR or IFR charts or the G1000 or a combination of the three to locate themselves. Some chose more effectively than others.

***Types of ADM tested.***

***In-flight ADM.*** The scenarios used to test the participants' ADM focused on in-flight ADM. Some pre-flight ADM was required, in that participants had to decide how thoroughly to review the planning information given to them, and how to manage their materials effectively in the AATD's pilot station. The extent to which the participants chose to review and manage the preflight planning information did impact the participants' performance, particularly when it came to locating his or her initial position and locating a diversion destination (if the scenario required it). However, the researcher

eliminated most of the pre-flight ADM by giving the participant the completed pre-flight planning documents. Therefore, weather information was given but the participant was not invited to verbalize a go or no-go decision. Likewise, weight and balance, performance calculations, and route and altitude choices were pre-determined for the participants.

*Quick decisions.* Several participants remarked that their particular actions were not “decisions.” When questioned further, some explained that they thought the action was too immediate or too simple to be called a decision. This suggested that many participants thought a decision must take a long time; another possible explanation is that the participants simply did not know how to explain their thought processes.

*Decision points in each scenario.* Table 6 shows the Overall ADM scores grouped by the scenario the participants flew in their first and second AATD sessions.

Table 6

*Overall ADM Scores and Delta Overall ADM, by Scenario Flown in Rounds One and Two*

Scenario Number	Round I	Round II	Delta
1	1.43	2.06	0.636
2	1.92	1.83	-0.083
3	1.84	1.83	-0.004
4	1.81	1.97	0.153

The four scenarios used to test the participants’ ADM included a variety of decisions. This variety meant that no one ADM process was the most appropriate to all of the scenarios. Generalizing the required ADM allowed the researchers to test for improvements in ADM without contamination from rehearsal effect. Further research

would be required to analyze the effectiveness of the SBT ADM training provided in this experiment for improving different types of decision-making.

The first decision that had to be made for each scenario was to determine how to fly the aircraft – what heading, altitude, and airspeed – and depended on the participants' ability to determine their location. In Scenario 1, this decision is the focus of the scenario. The participants were placed nearly 12 miles off shore over the Gulf of Mexico (due West of the Cross City VOR) at an altitude of 4,500 feet. This location was several miles off course, and put the participants beyond gliding distance of the shoreline in a single-engine aircraft. Had an engine failure occurred there, the participants would have had to ditch in open water.

The expected outcome was that the participants would see that they were over water, locate the nearest land, and turn immediately towards the nearest shoreline regardless of whether that course gave them a short intercept to the planned route of flight or necessitated back-tracking. The participants had the most difficulty with this scenario in the first AATD sessions, perhaps because they did not expect the problem to occur so soon in the scenario. ADM performance improved dramatically in the second AATD sessions, though, with participants scoring higher in Scenario 1 than any of the other scenarios (see Table 6).

Scenario 2 placed the participants on the planned route, 10 miles south of the Cross City VOR, on the planned heading of 352 degrees at 4,500 feet. This location was roughly mid-way between the departure airport and the destination. After reaching Cross City, the planned route turned to the northwest then due west. However, stronger than forecast winds (from the West at 60 knots) had been causing fuel to burn faster than

planned so that the participants started the scenario with only 17 gallons of fuel (the flight plan anticipated having 35 gallons remaining at that location). The fuel remaining would not have been enough to get them to the next planned waypoint after Cross City Airport given the winds and the minimum required fuel reserves.

The expected outcome was that the participants would decide to divert to a nearby airport such as Cross City Airport to refuel. In Scenario 2, the participants had more time to make a decision than they did in Scenario 1, depending on how quickly they noticed the fuel and the winds. Participants performed the best in this scenario in the first AATD sessions, but showed a slight regression in the second AATD sessions (see Table 6). This type of decision-making may have been more familiar to the participants before the experiment as current ERAU flight training emphasizes fuel management as part of cross-country flight planning.

A mechanical malfunction and light rain provided the main decision points in Scenario 3. Light rain began two minutes after the scenario started, although conditions remained VFR. Four minutes after the scenario started, the engine started running roughly. Engine roughness was indicated by variations in engine noise and RPMs indicated on the tachometer. The fluctuations in engine power worsened if the participant did not appear to notice them, and it persisted regardless of what the participants did to troubleshoot in the air. The expected outcome was that the participants would divert into a nearby airport before the engine failed or before encountering IFR conditions.

In Scenario 4, the participants had to decide what to do after their passenger started to get nauseous. The researchers simulated turbulent conditions starting two minutes into the scenario using the weather settings on the AATD. Then the researchers

role-played as the participant's "Auntie," getting progressively sicker and eventually vomiting. The participants were expected to divert to a nearby airport with the appropriate facilities to help the passenger recover.

*Unintended decisions.* The researcher assumed that the participants would be able to readily locate themselves with the resources available. This was not always the case, and difficulties locating themselves had a consequently negative impact on some participants' ADM with respect to the intended decision points. Other participants commented that they did not like being "dropped into the scenario in mid-air," because they felt it did not give them the time to prepare themselves or to set up the avionics the way they usually do. Others simply adapted to the lack of preparatory time, commenting that they used the bare minimum avionics to locate themselves and get on course before they took the time to set up the avionics the way they usually would.

*Maintaining realism of the scenario.* Conducting the scenarios in an AATD had many advantages. The AATD allowed the researcher to put the participant in scenarios that would be potentially life threatening in an aircraft without actually impacting the participants' safety. The AATD was also a more practical platform for training ADM since it was much more cost-effective to operate than a flight training device (FTD) or an aircraft, and it was already set up in a lab with video and audio recording equipment. Flying the AATD instead of an aircraft also allowed the researcher to control the variables of weather and maintenance.

However, using the AATD instead of an FTD or an aircraft introduced a challenge that is inherent in any simulation - maintaining the realism of the scenario. The researchers took many precautions to make the AATD sessions as life-like as possible –



to make the participants feel that they were in a real airplane dealing with that actual scenario. It was hoped that the participants would become involved enough that they would react as they would in a real airplane and gain equivalent experience. Several factors interfered with the desired realism and possibly reduced the effectiveness of the AATD sessions.

Several participants asked the researchers a question mid- scenario despite being briefed that the researchers would merely be observing once the scenario started. “Where am I?” was the most common question asked. Several participants also exhibited behaviors such as laughing when it was inappropriate to the scenario, indicating that they did not mentally accept the scenario as “real.” Other participants, confronted with simulated engine roughness, noted the fluctuations in RPM and engine noise but told the researcher during debrief that they thought it was “just the sim” and did nothing to address the engine roughness.

#### **Deviations from the scenario procedure.**

##### ***AATD-induced deviations from the scenario procedure.***

*Un-programmed engine failure.* The AATD induced unintended variations to several participants’ scenarios. On one occasion when the researcher programmed engine roughness, the AATD failed the engine instead. The researchers were unable to determine why the AATD failed the engine and allowed the simulation to continue, scoring that participant’s ADM based on the engine failure instead of the intended scenario. On multiple occasions, the AATD displayed cloud cover that was thicker than programmed by the researcher. The planned flight was to be conducted VFR and the researcher therefore programmed a low scattered layer that should not have been an obstacle to

descending VFR into an airport. However, several participants found themselves apparently in a solid cloud layer when they descended into what looked like a clear space between the scattered clouds on the visual display. This understandably changed those participants' ADM by introducing a VFR-into-IFR aspect to the scenario.

*Clouds and Visibility.* The researchers followed standardized procedures to set up the visual environment for each AATD session. The procedures specified VFR clouds and visibility for all four scenarios. One scenario called for rain, but the rain was not associated with lower visibility or cloud bases. In a few scenarios, the AATD displayed much thicker clouds than the researchers had programmed. The visibility also appeared significantly lower than programmed during one session. These aberrations caused several participants to request IFR clearance or to divert to an airport behind them.

*Auto-zoom.* Auto-zoom is a function of the G1000 that adjusts the zoom on the moving map display by referencing ground speed. The faster the ground speed, the farther out it zooms. An auto-zooming map can aid a pilot's SA by zooming out to show nearby airports and navigation aids. However, the researchers wanted to observe whether the participants would effectively use the moving map as a source of information. The researchers therefore intended to start everyone with the moving map zoomed in as tightly as possible and auto-zoom turned off. Auto-zoom activated and zoomed the range out on the moving map for some participants though, making it impossible to determine whether those participants would have zoomed the display out on their own or whether they would have tried to navigate without the moving map.

*Operator-induced deviations from the scenario procedure.* At two other times, the researchers inadvertently changed the scenario. In one instance, the researcher briefed

the participant enroute that the weather included a ceiling at 800 feet when the planned briefing called for a scattered cloud layer at 800 feet. That participant then requested an IFR clearance. A different participant started his scenario before the researchers realized that the fuel had been inadvertently left at the levels for the scenario given to the previous participant (which was low enough to require a fuel stop).

One participant's behavior during the second AATD session suggested that the participant probably talked with other study participants about the scenarios in their first AATD sessions. The scenario for that participant's second AATD session involved a sick passenger. The participant seemed to anticipate that the passenger would be sick, asking if she felt sick before the passenger had remarked on anything except the view. On one occasion, the researchers decided not to present a participant with all of the decisions called for by the scenario. The participant seemed to be overloaded already, and further challenges seemed unproductive.

**Treatment design factors.** Limitations of the study strongly influenced the design of the experimental treatment. Constraints included the participants' schedules, the study's condensed timeline, and budget. By necessity, the experimental treatment was very condensed, compared to other SBT ADM training efforts. The treatment consisted of one session lasting one and one-half hours. It was conducted in a classroom using a PowerPoint presentation as a visual aid, and some time elapsed before the participants completed their second AATD sessions.

Treatment would likely have been more effective if the resources had been available to conduct more than one session of ADM training. Harnessing participants' involvement in the material was limited to measures the researcher could take in the

treatment session. Increasing the participants' involvement by assigning homework or reading (for example, reading about ADM models or ADM-related accidents) beforehand could have helped them absorb the information. Also, the treatment would likely have been more effective if the guided discussion had been followed immediately by a hands-on application in a scenario in an AATD. Some participants waited several days after completing the treatment before schedules allowed them to complete the second AATD session.

**Researcher and instrument factors.**

*Bias.* The author conducted all of the AATD scenarios (with the help of another researcher), conducted the ADM training for the experimental group, and was the sole scorer for all of the participants. The fact that the same person conducted all aspects of the experiment reduced the researcher's impartiality. Knowing whether each participant was in the control or experimental group could have biased the researcher's scoring. The researcher also knew some of the participants through work as a flight instructor, others through student groups; this familiarity created another opportunity for bias to affect the results. Using multiple graders, who did not conduct any of the experimental training, to score the participants would have allowed inter-rater reliability to remove this bias.

*Scoring.* Quantifying a person's ADM is difficult by nature. The scoring method used in this experiment to quantify each participant's ADM was purposefully vague in order to lessen the impact of a single scorer's mistakes. The options of 1, 2, or 3 did not discriminate the participants' ADM as much as a 5-point or 7-point scale would have, but it allowed the scorer to reliably assess a participant's ADM. A 5-point or 7-point scale

would require very definitive criteria and several independent scorers to increase the validity and reliability of the assessments.

Each participant received several scores per AATD session. The researcher then averaged the scores for each participant to calculate an Overall ADM score for that participant in that session. This averaging process further smoothed out any scoring errors the researcher may have made.

*Instrument reliability.* The researcher conducted four Chronbach's alpha analyses to measure the research instrument's reliability. These analyses were especially important given some of the researcher's choices in designing the study. The study's use of a single person to train and score the participants created the possibility of bias and negated inter-rater reliability. The researcher also chose to use a relatively vague 3-point scale to score ADM. Furthermore, the literature review did not reveal a precedent for averaging scores for different aspects of ADM into an Overall ADM score.

The Chronbach's alpha analyses revealed that the instrument was extremely reliable despite the single scorer, vague scoring method, and unprecedented Overall ADM scores. Chronbach's alpha is measured on a scale of 0 to 1.0. An accepted "rule of thumb" for determining what is an acceptable alpha is:  $\alpha > 0.7$  acceptable,  $\alpha > 0.8$  good, and  $\alpha > 0.9$  excellent (George & Mallery, 2011, p. 231).

The Chronbach's alpha for the ordinal scores excluding Overall ADM in Round One and Round Two were excellent – 0.966 and 0.963 respectively. The Chronbach's alpha for the ordinal scores including Overall ADM in Round One and Round Two of 0.980 and 0.978 respectively were even higher. The researcher had expected the

Chronbach's alpha to be higher when the Overall ADM scores were included because the Overall ADM scores averaged scores that had already been shown to be highly reliable.

**Impact of video quality on data collection.** All AATD sessions were recorded using two cameras installed in the High Altitude Lab (HAL), headset microphones, and a portable audio recorder. The video recorded by the two cameras in the HAL was good enough to see what the participant was doing, but the poor lighting and resolution made it difficult or impossible to tell what kind of chart they were using or what any of the displays were showing. The cameras also did not record sounds that were not on the microphones. This means that audio cues like rain or changes in engine RPM were lost in the video recording. This made reviewing the footage more difficult. Some details that could have enabled the researcher to determine what the participant was thinking were lost in the video and could not be accurately recalled from the researcher's personal memories.

Reviewing the video and audio recordings after the participants completed the scenarios would have been easier, had the researcher modified the data collection device. The inclusion of a time log on the data collection device to record start and stop times and times of significant events would have facilitated observations both during the session and during later review. Recording the ground tracks would have facilitated evaluation particularly for the participants whose scenario started over water.

## **Conclusions**

The research hypothesis of this study was: there was a difference in demonstrated ADM ability between pilots who received scenario-based ADM training and pilots who

did not receive scenario-based ADM training, for flight students enrolled in a baccalaureate program at ERAU.

Statistical tests showed no significant difference between the control and experimental groups' ADM ability for any of the variables in the first AATD sessions. Although the experimental group exhibited higher mean scores than the control group for all variables in the first AATD sessions, the two groups exhibited statistically equivalent ADM ability. This meant that statistical differences in the two groups' ADM ability in the second AATD sessions did not result from an a priori difference.

The expected outcome of the treatment was that the experimental group would exhibit statistically better ADM ability than the control group in the second AATD sessions. Statistical tests showed no significant difference between the control and experimental groups' ADM ability in the second AATD sessions. Despite the experimental treatment, the ADM ability of the control and the experimental groups remained statistically equivalent.

The researcher expected both the control and the experimental groups to improve on their ADM scores from their first AATD sessions to their second AATD sessions. As expected, both the control and experimental groups showed higher mean scores for all variables in the second AATD session. Although both groups showed higher mean scores in the second AATD sessions, most of the differences were statistically insignificant.

The only statistically significant difference was in Problem Comprehended for the control group. Had alpha been set to 0.1 instead of 0.05, the differences in Projection ( $p = 0.085$ ) and Overall ADM ( $p = 0.089$ ) for the control group would have been significant. The difference in Problem Comprehended ( $p = 0.104$ ) for the experimental group would

have been close to significance. The general absence of statistically significant differences indicated that the experimental treatment had no significant impact on the participants' demonstrated ADM ability.

However, the treatment appeared to have a practically significant impact on the experimental groups' ADM. The experimental group's average Overall ADM improved from 1.82 in the first AATD session to 1.95 in the second AATD session (see Table 7). Although the experimental group's Delta (0.13) was smaller than the control group's Delta, the experimental group's Overall ADM scores were higher than the control group's Overall ADM scores in both the first and second AATD sessions (see Table 7).

Table 7

*Practical Comparison of the Control and Experimental Groups, Overall ADM*

	First AATD Session	Second AATD Session	Delta Overall ADM
Control	1.69	1.91	.21
Experimental	1.82	1.95	.13

The experimental group also scored higher than the control group in the second AATD session in Decision Process Used, Timely Manner, and Overall ADM.

It is possible that the size and the diversity of the population sample resulted in a Type II error. The population sample was small for an experiment of this nature ( $N = 30$ ). Although 30 subjects is a generally accepted minimum for statistical significance, using only the minimum number of participants and including participants with such a range of certifications and hours logged could have prevented a statistical difference in Overall ADM between the control and experimental groups.



## **Recommendations**

The data from this study did not statistically support incorporating SBT ADM training into GA flight training. However, previous research supports the hypotheses that ADM can be taught and that SBT is more effective than MBT for teaching ADM. The practical results of this study indicate that further research is warranted.

Recommendations from this study thus include recommendations for further research and suggested improvements to the experiment design.

**Further research and analysis.** Variations in the study's sample raised many questions about ADM training with respect to differences in total time and certifications held. However, subsets of the sample for different certifications and total times were too small to allow statistical analysis of those factors' impact on demonstrated ADM ability. Repeating this experiment with larger samples would allow meaningful analysis of the treatment's impact on ADM for GA pilots with different certifications and total times. Researchers could derive significant implications for effective ADM training for pilots at different experience levels and aid in development of graduated ADM training, which could then be integrated into GA flight training.

Further analysis of the participants' ADM scores by scenarios would enable researchers to derive implications on the effectiveness of ADM training for different types of decisions. This would require a sample of at least 30 participants for each scenario.

**Suggested improvements for this study.** This study had several limitations that significantly impacted the design of the experiment, the treatment design, the data analysis, and the results. Improvements for this study would address these limitations.

***Experiment design factors.*** Time and budget prevented the researcher from using a larger sample for the experiment. The researcher also used ERAU students exclusively because they were the most convenient. This study would benefit from using a larger sample from a more diverse GA pilot population. The experiment would also provide better results if the AATD sessions were recorded using cameras with higher quality video and audio. Poor lighting, resolution, and sound quality hindered the review of the AATD sessions and made scoring more difficult.

***Treatment design factors.*** Time and budget constrained the scope of the treatment the researcher could provide. The study would likely produce more conclusive results if the treatment were expanded beyond a single training session. The treatment would also be more effective if the researcher could integrate an AATD session into the end of treatment, as this would allow the participants to immediately practice the ADM processes they discussed in the treatment.

***Bias.*** The researcher conducted all of the AATD scenarios, conducted the training for the experimental group, and scored all of the participants. The researcher also knew some of the participants through work as a flight instructor and through student groups. Chronbach's alpha analyses indicated that bias did not impact the reliability of the data in this study. Using multiple raters (whose knowledge of the participants was limited to what they observed in the AATD sessions) to blindly score the participants would have allowed the use of inter-rater reliability to make the results even more reliable.

## References

- About SAFE*. (2012). Retrieved from <http://www.safepilots.org/about-safe/>
- Adams, R. J. (1992a). How expert pilots think. In Adams, R. J. & Adams, C. A., *Workshop on aeronautical decision making (ADM): Vol. II – plenary session with presentations and proposed action plan* (DOT/FAA/RD-92/14, II). Retrieved from <http://www.dtic.mil/dtic/tr/fulltext/u2/a257066.pdf>
- Adams, R. J. (1992b). Enhanced ADM training alternatives. In Adams, R. J. & Adams, C. A., *Workshop on aeronautical decision making (ADM): Vol. II – plenary session with presentations and proposed action plan* (DOT/FAA/RD-92/14, II). Retrieved from <http://www.dtic.mil/dtic/tr/fulltext/u2/a257066.pdf>
- Adams, R. J. & Adams, C. A. (1992). *Workshop on aeronautical decision making (ADM): Vol. II – plenary session with presentations and proposed action plan* (DOT/FAA/RD-92/14, II). Retrieved from <http://www.dtic.mil/dtic/tr/fulltext/u2/a257066.pdf>
- Adams, R. J., Hamilton, B. A., Koonce, J. M., & Hwoschinsky, P. V. (2002). Decision-making styles associated with accidents: Defining the high risk pilot. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 46, 948. doi: 10.1177/154193120204601203
- Adams, R., & Thompson, J. (1987). *Aeronautical decision making for helicopter pilots* (DOT/FAA/PM-86/45). Retrieved from [http://www.avhf.com/html/Library/ADM\\_for\\_Helicopter\\_Pilots.pdf](http://www.avhf.com/html/Library/ADM_for_Helicopter_Pilots.pdf)
- Aircraft Owners and Pilot's Association (AOPA). (2009). *2009 Nall report: Accident trends and factors for 2008*. Frederick, MD: Air Safety Institute.
- Aircraft Owners and Pilot's Association (AOPA). (2010). *2010 Nall report: The Joseph T. Nall report of accident trends and factors*. Frederick, MD: Air Safety Institute.
- Babbitt, J. R. (2011). Above and beyond: Attitude determines your altitude. *FAA Safety Briefing*, 50(4), 10-13.
- Baker, S. P., Qiang, Y., Rebok, G. W., & Li, G. (2008). Pilot error in air carrier mishaps: Longitudinal trends among 558 reports, 1983-2002. *Aviation Space Environmental Medicine*, 79(1), 2-6. doi: 10.3357/ASEM.2200.2008
- Beringer, D. B., & Schvaneveldt, R. (2002). Priorities of weather information in various phases of flight [Abstract]. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 46, 86-90. doi: 10.1177/154193120204600118

- Brown, M. W. (2006). System safety in modern flight training. *FAA Aviation News*, 45(1), 1-7.
- Buch, G., Lawton, R. S., & Livack, G. S. (Eds.) (1987). *Aeronautical decision making for instructor pilots* (DOT/FAA/PM-86/44). Washington, DC: Federal Aviation Administration (FAA). Retrieved from [http://www.avhf.com/html/Library/ADM\\_for\\_Instructor\\_Pilots.pdf](http://www.avhf.com/html/Library/ADM_for_Instructor_Pilots.pdf)
- Cessna Aircraft Company. (2005). *Information manual: Skyhawk SP Model 172S Nav III avionics option*. Wichita, KS: Author.
- Center for General Aviation Research (CGAR). (2005). *CGAR 2.1*. Retrieved from <http://www.cgar.org/Default.aspx>
- Center for General Aviation Research (CGAR). (2011). *10<sup>th</sup> anniversary annual report*. Retrieved from <http://www.cgar.org/newsletters/Annual%20Report%202011.pdf>
- Deitch, E. L. (2001). *Learning to land: A qualitative examination of pre-flight and in-flight decision-making processes in expert and novice aviators* (Unpublished dissertation). Virginia Polytechnic Institute and State University, Blacksburg, Virginia. URN: etd-12122001-183807
- Diehl, A. E. (1992). Does cockpit management training reduce aircrew error? In Adams, R. J. & Adams, C. A., *Workshop on aeronautical decision making (ADM): Vol. II – plenary session with presentations and proposed action plan* (DOT/FAA/RD-92/14, II). Retrieved from <http://www.dtic.mil/dtic/tr/fulltext/u2/a257066.pdf>
- Diehl, A. E., Hwoschinsky, P. V., Lawton, R. S., & Livack, G. S. (Eds.) (1987). *Aeronautical decision making for student and private pilots* (DOT/FAA/PM-86/41). Washington, DC: Federal Aviation Administration (FAA). Retrieved from [http://www.avhf.com/html/Library/ADM\\_for\\_Student\\_and\\_Private\\_Pilots.pdf](http://www.avhf.com/html/Library/ADM_for_Student_and_Private_Pilots.pdf)
- Dillman, B. G., & Fanjoy, J. R. (2006). Utilizing situational judgment tests (SJT) for pilot decision-making. *International Journal of Applied Aviation Studies*, 6(1), 145-154. Retrieved from [http://www.faa.gov/about/office\\_org/headquarters\\_offices/arc/programs/academy/journal/pdf/Spring\\_2006.pdf](http://www.faa.gov/about/office_org/headquarters_offices/arc/programs/academy/journal/pdf/Spring_2006.pdf)
- Endsley, M. R. (2000). Theoretical underpinnings of situational awareness: A critical review. In Endsley, M. R. & Garland, D. J. (Eds.) *Situation awareness analysis and measurement*. Mahwah, NJ: Lawrence Erlbaum Associates. doi: 10.1177/154193120204601203
- Endsley, M. R., & Garland, D. J. (2000). Pilot situation awareness training in general aviation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 44(11), 357-360. doi: 10.1177/154193120004401107

- Enrollment, ERAU-News, Embry-Riddle.* (2012). Retrieved from <http://news.erau.edu/media-resources/facts-figures/enrollment/index.html>
- Ericsson, K. A. (1992). Methodology for studying and training expertise. In Adams, R. J. & Adams, C. A., *Workshop on aeronautical decision making (ADM): Vol. II – plenary session with presentations and proposed action plan* (DOT/FAA/RD-92/14, II). Retrieved from <http://www.dtic.mil/dtic/tr/fulltext/u2/a257066.pdf>
- FAA-industry training standards (FITS) program plan.* (2003). Retrieved from [http://www.faa.gov/training\\_testing/training/fits/](http://www.faa.gov/training_testing/training/fits/)
- Federal Aviation Administration (FAA). (1991). *Aeronautical decision making* (AC 60-22). Retrieved from [http://www.faa.gov/regulations\\_policies/advisory\\_circulars/index.cfm/go/document.information/documentID/22624](http://www.faa.gov/regulations_policies/advisory_circulars/index.cfm/go/document.information/documentID/22624)
- Federal Aviation Administration (FAA). (2008a). *Aviation instructor's handbook* (FAA-H-8083-9A). Washington, DC: Author.
- Federal Aviation Administration (FAA). (2008b). *Pilot's handbook of aeronautical knowledge* (FAA-H-8083-25A). Washington, DC: Author.
- Federal Aviation Administration (FAA). (2010a) *Flight instructor instrument practical test standards for airplane and helicopter* (FAA-S-8081-9D with change 1). Washington, DC: Author.
- Federal Aviation Administration (FAA). (2010b). *Instrument rating practical test standards for airplane, helicopter, and powered lift* (FAA-S-8081-4E with changes 1, 2, and 3). Washington, DC: Author.
- Federal Aviation Administration (FAA). (2011a). *Commercial pilot practical test standards for airplane (SEL, MEL, SES, MES)* (FAA-S-8081-12C). Washington, DC: Author.
- Federal Aviation Administration (FAA). (2011b). *Private pilot practical test standards for airplane (SEL, MEL, SES, MES)* (FAA-S-8081-14B). Washington, DC: Author.
- FITS master instructor syllabus: TAA scenario based instructor guide* (version 2.0). (2006). Retrieved from [http://www.faa.gov/training\\_testing/training/fits/training/generic/media/instructor.pdf](http://www.faa.gov/training_testing/training/fits/training/generic/media/instructor.pdf)
- Frasca International, Inc. (2010). *Frasca International - products - fixed wing piston - Mentor G1000* [sic]. Retrieved from <http://www.frasca.com/products/mentorg100.php#>

- French, J., Blickensderfer, B., Ayers, F., & Connolly, T. (2005). *FITS combined task 1&2 final report for the Embry Riddle Aeronautical University effort*. Retrieved from [http://www.faa.gov/training\\_testing/training/fits/research/media/erau.pdf](http://www.faa.gov/training_testing/training/fits/research/media/erau.pdf)
- Garmin G1000®. (2012). Retrieved from <https://buy.garmin.com/shop/shop.do?pID=6420&ra=true>
- General Aviation Manufacturers Association (GAMA). (2011). *General aviation: Statistical databook and industry outlook*. Retrieved from <http://www.gama.aero/media-center/industry-facts-and-statistics/statistical-databook-and-industry-outlook>
- George, D., & Mallery, P. (2011). *SPSS for Windows step by step: A simple guide and reference 18.0 update* (11<sup>th</sup> ed.). Boston, MA: Allyn & Bacon.
- Glista, T. (2003). FITS-FAA/industry training standards: Part 1 overview. *FAA Aviation News*, 42(2), 1-4.
- Helmreich, R. L., Merritt, A. C., & Wilhelm, J. A. (1999). The evolution of Crew Resource Management training in commercial aviation. *International Journal of Aviation Psychology*, 9(1), 19-32. Retrieved from <http://homepage.psy.utexas.edu/homepage/group/helmreichlab/publications/pubfiles/Pub235.pdf>
- Helmreich, R. L., & Foushee, C. (2010). Why CRM? Empirical and theoretical bases for human factors training. In Kanki, B. G., Helmreich, R. L., & Anca, J. (Eds.), *Crew resource management* (2<sup>nd</sup> ed.). Burlington, MA, USA: Academic Press.
- Howell, J., Miller, P., Park, H. L., Sattler, D., Schack, T., Sperry, E., et al. (2005). *Reliability and validity: Writing@CSU*. Retrieved from the Colorado State University Department of English website: <http://writing.colostate.edu/guides/research/relval/>
- Hunter, D. R. (2003). Measuring general aviation pilot judgment using a situational judgment technique. *The International Journal of Aviation Psychology*, 13(4), 373-386. Retrieved from [http://www.avhf.com/html/Library/Tech\\_Reports/IJAP\\_Situational\\_Judgment.pdf](http://www.avhf.com/html/Library/Tech_Reports/IJAP_Situational_Judgment.pdf)
- International Civil Aviation Organization. (2009). *Safety management manual* (2nd ed.). Retrieved from [http://www2.icao.int/en/ism/Guidance%20Materials/DOC\\_9859\\_FULL\\_EN.pdf](http://www2.icao.int/en/ism/Guidance%20Materials/DOC_9859_FULL_EN.pdf)

- Irving, J. (1992). Bootstrapping expertise: What makes the expert pilot decision maker expert and why? In Adams, R. J. & Adams, C. A., *Workshop on aeronautical decision making (ADM): Vol. II – plenary session with presentations and proposed action plan (DOT/FAA/RD-92/14, II)*. Retrieved from <http://www.dtic.mil/dtic/tr/fulltext/u2/a257066.pdf>
- Jensen, R. S. (1982). Pilot judgment: Training and evaluation. *Human Factors*, 24(1), 61-73. doi: 10.1177/001872088202400107
- Jensen, R. S. (1989). *Aeronautical decision making – Cockpit resource management*. (DOT/FAA/PM-86/46). Retrieved from [http://www.avhf.com/html/Library/Aeronautical\\_Decision\\_Making\\_Cockpit\\_Resource\\_Management.pdf](http://www.avhf.com/html/Library/Aeronautical_Decision_Making_Cockpit_Resource_Management.pdf)
- Jensen, R. S., & Adrion, J. (1988). *Aeronautical decision making for commercial pilots* (DOT/FAA/PM-86/42). Washington, DC: Federal Aviation Administration (FAA). Retrieved from [http://www.avhf.com/html/Library/ADM\\_for\\_Commercial\\_Pilots.pdf](http://www.avhf.com/html/Library/ADM_for_Commercial_Pilots.pdf)
- Jensen, R. S., Adrion, J., & Lawton, R. S. (1987). *Aeronautical decision making for instrument pilots* (DOT/FAA/PM-86/43). Washington, DC: Federal Aviation Administration (FAA). Retrieved from [http://www.avhf.com/html/library/ADM\\_for\\_Instrument\\_Pilots.pdf](http://www.avhf.com/html/library/ADM_for_Instrument_Pilots.pdf)
- Kobus, D. A., Proctor, S., Bank, T., & Holste, S. (2000). Effects of experience and uncertainty during dynamic decision making. *Proceedings of the Human Factors and Ergonomics Society*.
- Kochan, J. A., Jensen, R. S., Chubb, G. P., & Hunter, D. R. (1997). *A new approach to aeronautical decision-making: The expertise method* (DOT/FAA/AM-97/6). Washington, DC: Federal Aviation Administration.
- Lee, J. R., Fanjoy, R. O., & Dillman, B. G. (2005). Sorenson best paper award recipient: The effects of safety information on aeronautical decision making. *Journal of Air Transportation*, 10(3).
- Li, W.-C. & Harris, D. (2005). Aeronautical decision making: Instructor-pilot evaluation of five mnemonic methods. *Aviation, Space, and Environment Medicine* 76(12), 1156 – 1161.
- Li, W.-C. & Harris, D. (2008). The evaluation of the effect of a short aeronautical decision-making training program for military pilots. *The International Journal of Aviation Psychology*, 18(2), 135-152.
- McMahon, A. (2009). Making it fun: Scenario-based training is a great and effective way to train today's pilots. *FAA Aviation News*, 48(4), 8-11.

- Molesworth, B., Wiggins, M. W., & O'Hare, D. (2006). Improving pilots' risk assessment skills in low-flying operations: The role of feedback and experience. *Accident Analysis and Prevention*, 38, 954-960.
- National Transportation Safety Board (NTSB). (2006). *Annual review of aircraft accident data: U.S. general aviation, calendar year 2006* (NTSB/ARG-10/01). Washington, DC: Author.
- National Transportation Safety Board (NTSB). (2011a). *Review of U.S. civil aviation accidents, 2007–2009. Annual Review* (NTSB/ARA-11/01). Washington, DC: Author.
- National Transportation Safety Board (NTSB). (2011b). *Table 5: Accidents, fatalities, and rates, 1991-2010, 14 CFR 121, scheduled service (airlines)*. Retrieved from [http://www.nts.gov/data/aviation\\_stats.html](http://www.nts.gov/data/aviation_stats.html)
- National Transportation Safety Board (NTSB). (2011c). *Transportation safety: General Aviation safety*. <http://www.nts.gov/safety/mwl-2.html>
- National Transportation Safety Board (NTSB). (2012). *General aviation safety: Climbing to the next level*. Retrieved from [http://www.nts.gov/news/events/2012/GA\\_safety/index.html](http://www.nts.gov/news/events/2012/GA_safety/index.html)
- O'Brien, K., & Ballenger, K. (2006). Introducing the FAA Safety Team. *FAA Aviation News* 45(2), 1-4.
- O'Connor, P., Hahn, R. G., & Nullmeyer, R. (2010). The military perspective. In Kanki, B. G., Helmreich, R. L., & Anca, J. *Crew resource management* (2<sup>nd</sup> edition). Burlington, MA, USA: Academic Press.
- O'Hare, D., Mullen, N., & Arnold, A. (2009). Enhancing aeronautical decision making through case-based reflection. *The International Journal of Aviation Psychology*, 20(1), 48-58.
- Robertson, C. L., Petros, T. V., Schumacher, P. M., McHorse, C. A., & Ulrich, J. M. (2006). *Evaluating the effectiveness of FITS training*. Retrieved from [http://www.faa.gov/training\\_testing/training/fits/research/media/und.pdf](http://www.faa.gov/training_testing/training/fits/research/media/und.pdf)
- Robertson, C. L., & Summers, M. (2007). *Evaluating the satisfaction and quality of FITS flight training*. Retrieved from [http://www.faa.gov/training\\_testing/training/fits/research/media/satisfaction.pdf](http://www.faa.gov/training_testing/training/fits/research/media/satisfaction.pdf)
- Schriver, A. T., Morrow, D. G., Wickens, C. D., & Talleur, D. A. (2008). Expertise differences in attentional strategies related to pilot decision making. *Human Factors*, 50(6), 864-878. doi: 10.1518/001872008X374974



- Society of Aviation and Flight Educators (SAFE). (2011a). *Pilot training reform symposium preliminary report*. Retrieved from [http://www.pilottrainingreform.org/documents/Symposium\\_Prelim\\_Report\\_06Jun2011.pdf](http://www.pilottrainingreform.org/documents/Symposium_Prelim_Report_06Jun2011.pdf)
- Society of Aviation and Flight Educators (SAFE). (2011b). *Update on training reform symposium: Preliminary report and stakeholder comments*. Retrieved from [http://www.pilottrainingreform.org/documents/SAFE\\_Training\\_Reform\\_Progress\\_Report\\_31Oct11.pdf](http://www.pilottrainingreform.org/documents/SAFE_Training_Reform_Progress_Report_31Oct11.pdf)
- SRM Scenario [PowerPoint slides]. (2009). Retrieved from [http://www.faa.gov/training\\_testing/training/fits/presentations/](http://www.faa.gov/training_testing/training/fits/presentations/)
- Stowell, R. (2011). *Symposium acts as springboard for training reform*. Retrieved from <http://www.pilottrainingreform.org/2011/05/symposium-acts-as-springboard-for-training-reform/>
- Student Demographics, ERAU-News, Embry-Riddle*. (2012). Retrieved from <http://news.erau.edu/media-resources/facts-figures/student-demographics/index.html>
- Sumwalt, R. L., & Watson, A. W. (1995). ASRS incident data reveal details of flight-crew performance during aircraft malfunctions. *Flight Safety Digest* 14(10), 1-7.
- Vision & mission statement*. (2012). Retrieved from <http://www.safepilots.org/about-safe/vision-mission-statements/>
- Wiegmann, D. A., Goh, J., & O'Hare, D. (2002). The role of situation assessment and flight experience in pilots' decisions to continue visual flight rules flight into adverse weather. *Human Factors*, 44(2), 189-197. doi: 10.1518/0018720024497871
- Wiggins, M., & O'Hare, D. (2003). Weatherwise: Evaluation of a cue-based training approach for the recognition of deteriorating weather conditions during flight. *Human Factors*, 45(2), 337-345. doi: 10.1518/hfes.45.2.337.27246
- Wright, R. A. (2002). *White paper: Changes in general aviation flight operations and their impact on system safety and flight training*. Retrieved from [http://www.faa.gov/training\\_testing/training/fits/guidance/media/whitepaper.doc](http://www.faa.gov/training_testing/training/fits/guidance/media/whitepaper.doc)
- Wright, R. A. (2004). General aviation safety in the second century: The flight training challenge. *FAA Aviation News*, 43(4), 4-5.
- Wright, R. A. (2009). *The current state of flight instruction and flight training in the United States: A call for industry action*. Retrieved from [http://www.pilottrainingreform.org/documents/PTR\\_WP\\_Wright.pdf](http://www.pilottrainingreform.org/documents/PTR_WP_Wright.pdf)

Wright, R. A. (2011). Training's future: Replacing the FITS program means improving and refining its emphasis on enhanced decision making and automation management. *Aviation Safety* 31(1), 8-11.

**Appendix A**

**Permission to Conduct Research**

**Embry-Riddle Aeronautical University****Application for IRB Approval****Determination Form****11-167**

Principle Investigator: Dr. Guy Smith

Other Investigators: Mariko Doskow, Michele Halleran, Michael Wiggins

Project Title: *Classroom and lab-based experiment to support a thesis*Submission Date: October 2, 2011Determination Date: October 28, 2011Review Board Use Only

---

Initial Reviewer: Teri Vigneau/Bert BoquetExempt:  Yes  NoApproved:  Yes  No

Comments: The purpose of this project is to measure the effectiveness of scenario-based training for accelerating improvement of judgment. There will be one control group with no ADM (Aeronautical Decision Making) training and an experimental group with ADM training.

Part of the study will be conducted in a classroom and part will take place in an AATD (Mentor Advanced) fixed (non-moving) flight simulator. There should be no risk to participants other than routine training so it may be considered **exempt**. [Teri Vigneau 10-24-11]

This falls under the university's scope of operations and can be considered exempt. [Bert Boquet 10-24-11]

**Appendix B**  
**Participant Consent Form**

**CONSENT FORM**  
**Embry-Riddle Aeronautical University**

I consent to participating in the research project entitled:

**Improving Aeronautical Decision Making Ability through Specialized Training**

The principle investigator of the study is:

**Dr. Guy Smith**

**Mariko Doskow, First student investigator**

**Prof. Michele Halleran, Advisor**

**Dr. Michael Wiggins, Advisor**

The purpose of this study is to measure pilot judgment. The participants will complete two separate scenarios in a Mentor AATD, a fixed-based (non-moving) flight simulator. Some participants will participate in an additional classroom session. Participants must have at least a student pilot certificate. Risks associated with participation are comparable to ERAU training in a fixed base simulator or classroom. Potential benefits include a valuable learning experience and an input for a student's resume.

Total time commitment for participants will total between 3 and 4.5 hours. Participants will be paid at a rate of \$7.50 per hour. Participants have the right to refuse participation at any time with no penalty or prejudice against them; however, participants who do not complete all portions of the study (regardless of whether they are asked to participate in two or all three sessions) will not be compensated. All personal information and experimental data collected for this study will be kept confidential.

The individual above, or their research assistants, have explained the purpose of the study, the procedures to be followed, and the expected duration of my participation. Possible benefits of the study have been described, as have alternative procedures, if such procedures are applicable and available.

I acknowledge that I have had the opportunity to obtain additional information regarding the study and that any questions I have raised have been answered to my full satisfaction. Furthermore, I understand that I am free to withdraw consent at any time and to discontinue participation in the study without prejudice to me.

Finally, I acknowledge that I have read and fully understand the consent form. I sign it freely and voluntarily. A copy has been given to me.

Date: \_\_\_\_\_

Name (*please print*): \_\_\_\_\_  
 (Participant)

Signed: \_\_\_\_\_  
 (Participant)

Signed: \_\_\_\_\_  
 (Researcher/Assistant)

**Appendix C**

**Pilot Briefing**

## **PILOT BRIEFING**

Thank you for volunteering to be a test pilot for this research project! The purpose of this study is to measure pilot judgment. The participants will complete two separate scenarios in a Mentor AATD, a fixed-based (non-moving) flight simulator. This Mentor is similar to the ones used for ADF training during ERAU's instrument training course. Please keep in mind that we are not grading you – we are collecting data that will be de-identified and used to make future training improvements.

We know that you will be tempted to tell your friends and colleagues about your experience. We ask that you refrain from discussing what you do, though. It is critical that all participants begin each scenario without any extra information in order to draw valid conclusions from this research.

Each scenario begins in level cruise enroute on the VFR cross country flight detailed in the flight plan and documents provided (KRSW to KHRT, departing Mon Oct 31, 2011). We hope that you have fun during this session. Fly safe and enjoy the challenge!



**Appendix D**  
**Sample Scoring Sheet**

Name	Scenario #	Decision Points	Round 1:	Round 1:	Round 1:	Round 1:	Round 1:	Round 1:	Round 1:	Round 1:	Round 1:	Round 1:
			Detected	Comprehension	Projection	Decision	Resolved	Timely	Safe	Overall		
			Y/N	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	Y/N	1, 2, 3	Y/N	1, 2, 3	Y/N
Participant 1	1	<b>Scenario Average:</b> Where am I? How should I fly now? How should I get back w/n gliding distance? What route should I take from there?	Y	1.667	1.667	1.667	1.667	2 Y	2.000	2 N	2 N	1.750
Participant 2	2	<b>Scenario Average:</b> Where am I? How should I fly now? How should I react to the crosswind/headwind? Fuel is too low - what should I do?	Y	3.000	3.000	3.000	3.000	3 Y	3.000	3 Y	3 Y	3.000
Participant 3	3	<b>Scenario Average:</b> Where am I? How should I fly now? Rain starts - should I change my plan? Engine is running rough - what should I do?	Y	2.000	2.333	2.333	2.000	2 Y	2.333	2 Y	2 Y	2.167
Participant 4	4	<b>Scenario Average:</b> Where am I? How should I fly now? Turbulence starts - should I change my plan? Passenger is sick - what should I do now?	Y	2.000	2.000	2.000	1.500	2 Y	2.000	2 Y	2 Y	1.875

**Appendix E**  
**Sample Debrief Form**

### Guided Discussion/Debrief

Date:

Scenario:

Total time:

1. What was your first concern? Next concern?
  
2. What was your first action?
  
3. Did you feel that you had any decisions to make?
  
4. What choices were you considering?
  
5. What resources did you use to help decision making?
  
6. What were your choices in the end?
  
7. What was your final decision? Why?
  
8. What concerns did you have after you chose to ...?

**Appendix F**

**Sample AATD Session Procedure**

Scenario 2 – Strong Westerly Headwinds Late in a Cross Country  
AATD Setup Procedure

1. Complete AATD Startup Checklist.
2. Wait for the communication channel to be reached, the screen will change colors until arriving at the default runway 7L DAB
3. Follow 172 setup checklists to initiate glass cockpit, ensure cockpit controls are set up for flight and ready for the scenario to begin.
 

Flaps – Up	Standby Battery – On
Mixture – Rich	Ignition – Both
Throttle – 2400 RPM	Parking Brake – In
Trim – Neutral	Standby Static Source – In
Electric Switches – Off	Fuel Shutoff – In
Master Switch – On	Fuel Selector – Both
Avionics Switch – On	FREEZE Red Button – On
	On MFD – Press ENTER
4. Setup scenario on Gist laptop:
  - a. Initial environment:
    - i. Cloud Layer 1: Bkn 300-1,500'. Cloud Layer 2: Sct 7,000-10,000'
    - ii. Wind @ Sfc, 2,000': 270@20. Wind @ 4,500': 270@60
    - iii. Day
    - iv. VIS ON @ 20sm, and Scud OFF
  - b. Initial fuel on board: 17gallons
  - c. Initial position: KAJYE (66 NM north of PIE/37 south of CTY on V35)
  - d. Initial attitude:
    - i. Wings level, pitch +2\*
    - ii. Heading 352\*, altitude 4,500', Airspeed 110 kts
5. Settle the subject in the simulator chair. Have the subject start the scenario.
  - a. After subject is seated, start recording
    - i. Start cameras
    - ii. Start audio recorder
  - b. “When you’re ready, go ahead and start the scenario by pressing the red pause/unpause button”
6. Scenario Timeline:
  - a. Subject starts the scenario; no further action by the HAL operator.
  - a. Subject notices where they are, low fuel and GS (Problem detected)
  - b. Subject projects whether or not can make the destination
  - c. Participant starts decision process: divert, checklists, communicate
  - d. After on a course for :03 (or on original course for :30), tell the participant to freeze the Mentor. “Please press the red pause/unpause button.”

7. Complete the Guided Discussion. “Thank you for your time. You will be compensated for this session (1 hour) after completing all sessions.”
8. Repeat steps 3-4 to reset the simulator controls to flight ready conditions.
9. Repeat steps 5-7 for next subject.

When the testing day is concluded, complete AATD shutdown checklist